THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXIX NOVEMBER 1960 NUMBER 6

Signaling Systems for Control of Telephone Switching
C. BREEN AND C. A. DAHLBOM 1381

The E6 Negative Impedance Repeater
A. L. BONNER, J. L. GARRISON AND W. J. KOPP 1445

Interstitial Channels for Doubling TD-2 Radio System Capacity
H. E. CURTIS, T. R. D. COLLINS AND B. C. JAMISON 1505

A Study of Talking Distance and Related Parameters in Hands-Free Telephony
M. B. GARDNER 1529

Magnetic Latching Relays Using Glass-Sealed Contacts
P. HUSTA AND G. E. PERREAULT 1553

Potential Distribution and Capacitance of a Graded p-n Junction
S. P. MORGAN AND F. M. SMITS 1573

Transverse Electron Beam Waves in Varying Magnetic Fields
E. I. GORDON 1603

"Ionic Radii," Spin-Orbit Coupling and the Geometrical Stability of Inorganic Complexes
A. D. LIEHR 1617

Molecular Structure in Crystal Aggregates of Linear Polyethylene
R. D. BURBANK 1627

Recent Bell System Monographs 1665

Contributors to This Issue 1670

COPYRIGHT 1960 AMERICAN TELEPHONE AND TELEGRAPH COMPANY

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

H. I. ROMNES, *President, Western Electric Company*

J. B. FISK, *President, Bell Telephone Laboratories*

E. J. McNEELY, *Executive Vice President, American Telephone and Telegraph Company*

EDITORIAL COMMITTEE

A. C. DICKIESON, *Chairman*
S. E. BRILLHART
A. J. BUSCH
L. R. COOK
R. L. DIETZOLD
J. H. FELKER
K. E. GOULD
E. I. GREEN
G. GRISWOLD, JR.
J. R. PIERCE
M. SPARKS
W. O. TURNER

EDITORIAL STAFF

W. D. BULLOCH, *Editor*
R. M. FOSTER, JR., *Assistant Editor*
C. POLOGE, *Production Editor*
J. T. MYSAK, *Technical Illustrations*
T. N. POPE, *Circulation Manager*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. F. R. Kappel, President; S. Whitney Landon, Secretary; L. Chester May, Treasurer. Subscriptions are accepted at $5.00 per year. Single copies $1.25 each. Foreign postage is $1.08 per year or 18 cents per copy. Printed in U.S.A.

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXXIX NOVEMBER 1960 NUMBER 6

Copyright 1960, American Telephone and Telegraph Company

Signaling Systems for Control of Telephone Switching

By C. BREEN and C. A. DAHLBOM

(Manuscript received May 17, 1960)

Telephone signaling is basically a matter of transferring information between machines, and between humans and machines. The techniques developed to accomplish this have evolved over the years in step with advances in the total telephone art. The history of this evolution is traced, starting from the early simple manual switchboard days to the present Direct Distance Dialing era. The effect of the increasing sophistication in automatic switching and transmission systems and their influence on signaling principles are discussed. Emphasis is given to the signaling systems used between central offices of the nationwide telephone network and the influence on such systems of the characteristics of switching systems and their information requirements, the transmission media and the compatibility problem. A review is made of the forms and characteristics of some of the interoffice signaling systems presently in use. In addition, the problem of signaling between Bell System and overseas telephone systems is reviewed with reference to differing information requirements, signaling techniques and new transmission media. Finally, some speculation is made on the future trends of telephone signaling systems.

I. INTRODUCTION

Telephony in the United States has now reached the point where it is becoming commonplace for an increasing number of customers to directly dial their calls throughout the nation. By the end of 1959 more

than 15 million main telephones in the Bell Telephone Companies had access to nationwide dialing, and eventually many more millions of telephone users will be able to quickly and accurately communicate with each other in the same way.

To a large degree this modern "miracle" has been made possible not only by the development of the common control switching system but also by advances in the art of signaling. Just as the switching function has sometimes been called the "giant brain", the signaling function might indeed be thought of as a gigantic "nerve system" carrying sensing and action impulses to and from the "brain" to every part of the system, near and far.

Today's communication system, which lets millions of telephone users throughout the nation get in touch with each other within seconds after dialing a number, has resulted from the effort of imaginative telephone people working over the years to evolve new techniques to meet the ever growing communication needs of our nation. The present highly developed state of the art has been brought about by an orderly evolution — an evolution wherein communication needs, plans and techniques, the one stimulating the other, have formed a continuously improving technology. Some appreciation of the place of the signaling art in the modern telephone system can be gained by reviewing this evolution.

II. HISTORICAL REVIEW

In the simple early telephone system, customers needing only to attract the attention of an operator at a switchboard cranked the familiar hand magneto, and thereby immediately brought the signaling function into play. In handling the call thereafter, many of the operator functions involved the concept of signaling. Such functions as becoming aware of calls in various stages of completion, obtaining verbal instructions as to the destination of the call, determining whether the called line is busy or idle, alerting the called station and reporting on the status of the call reduce to information signals of one form or another either to or from the operator. As such, they were either aids to, or results of, the judgment exercised in the operator's handling of calls.

Rather early in the evolution of the telephone it became apparent that the judgment and efficiency of the operator could be helped by improving the signals she was called upon to interpret and produce. Thus, the "drop" gave way to the line lamp, recall and disconnect

signals were displayed on lamps associated with the switchboard cords and signals produced by the operator to advance the call to an intermediate operator or to complete it to the called customer tended to be produced automatically. However, as long as manual switchboard operation prevailed, one part of the call handling process remained relatively unchanged — the verbal transmission of the called number from customer to operator, and from operator to operator.

With the advent of automatic or machine switching it became necessary for people to communicate with machines, for machines to communicate with other machines, and for machines to communicate with people. Machine simulation of many of the signals was accomplished more or less readily since, in practically all cases, the information items could be reduced to two-state signals. For example, either a line is requesting service or it is not; either it is busy or it is not; either it is being rung or it is not; either the telephone is removed from the switchhook or it is "on-hook", etc.

The adequacy of two-state signaling systems for these functions was apparent from the start and has had an important bearing on the subsequent evolution of signaling systems. However, the "addressing" signals, that is, the signals needed for people to tell machines to whom they wish to be connected and for machines to communicate this information to other machines, have a much larger information content than do the signals used for status or control purposes. Consequently, the handling of the "address" function has produced some highly interesting variations during the evolution of telephone signaling systems.

At the beginning, systems used for the automatic transmission of addressing information were based on the use of direct-current electrical pulses, the number of which denoted each decimal digit. Thus, a customer's telephone "address" could be designated by one or more decimal digits and these, in the form of pulses, could be transmitted to, and interpreted by, the switching machine. Of interest is the fact that in some of the earliest systems such pulses were hand-pulsed by the calling party with a pushbutton, different pushbuttons being used for transmitting sequential digits.

A major advance in the art occurred after the turn of the century with the introduction of the dial, a device which rotates back to a stop from a position to which it is pulled, and, in so doing, automatically generates a train of pulses corresponding to the value of the desired decimal digit. These pulsed digits directly positioned the switches. The step-by-step switching system thus evolved from this method of pulsing and the address information — from people to machines and

machines to machines — was conveyed in all cases by pulsed decimal digits.

Panel and crossbar switching systems, which were introduced later, utilized switching mechanisms which did not directly respond to, or follow, the calling subscriber pulses. These systems required auxiliary pulse-receiving and pulse-sending arrangements, known as senders, which received digital information independently of the switch and then controlled the position of the switch in a manner best suited to that device and to the system. The use of these auxiliary signaling arrangements permitted alternative actions by the machine in controlling the switches. The introduction of such arrangements in panel, and later crossbar, systems also had the effect of permitting information storage and signal regeneration in the switching office, thereby opening up possibilities for new interoffice "languages" for transmitting the address information. Thus were born the signaling systems which we now know as revertive, panel call indicator (PCI) and multifrequency. To more effectively utilize the speed capabilities of the sender-type devices, these signaling systems employed multistate signaling techniques.

The transformation to machine or dial central office operation occurred first in local exchange areas, and, since distances were relatively short, DC signal transmission was adequate for the signaling languages used by step-by-step, revertive and PCI systems. The later expansion of machine operation to longer distances, and the handling of toll traffic, accented the need for accuracy and speed, and also for signaling arrangements adaptable to non-DC paths. Out of this grew the extensive use of multifrequency pulsing for the transmission of digital switching instructions. Also appearing on the scene were single-frequency signaling systems, for transmitting trunk control and status information, and for transmitting dial pulses where required by the older switching systems. Single-frequency and multifrequency systems have become the predominant standard for modern exchange and toll applications where transmission of switching and supervisory information is required over AC paths. In fact, transmission of digital switching information by multifrequency pulses not only is used over AC paths but also has become the preferred method of signaling over all paths in toll and exchange switching systems.

These, then, are some of the principal events highlighting the history of signaling in the telephone system. Those close to the art will recognize that much has been left unsaid. We shall forego further detail at this point, however, in favor of an examination of the philosophy and principles underlying the evolution of the signaling art.

III. PRINCIPLES OF SIGNALING SYSTEMS

A telephone system, in simplest terms, consists of a network wherein any customer's telephone station can be used to signal, connect to and talk with any other similar station in the network. Within the framework of this basic system function, modern telephone system design is aimed at meeting numerous objectives — automation, high-grade transmission, accuracy, speed, flexibility are some of the principal ones, not forgetting the all-important objective of engineering economy. The signaling philosophy has evolved during the process of integrating the signaling function with the basic system function, while observing all of the subordinate system objectives.

Modern signaling systems are closely related to common control switching systems. From the viewpoint of the signaling engineer, the common control system appears as a machine receiving external stimuli in the form of instructions and other controls, following which it performs its required switching operations and generates additional information or instructions. Some of the information generated is returned to the source and some, together with original information, is transmitted to an external destination. It is the function of a signaling system to accept source information, convert it to appropriate signals for transmission over telephone lines or trunks, and deliver it to its destination.

The signaling function begins with the acceptance of input information from a source, followed by processes of encoding, signal generation, signal transmission, signal detection, decoding and delivery of information to a destination. While these processes may be combined in varying degrees in different signaling systems, they are nevertheless present in all.

The source and destination of the information transmitted or received by any particular signaling link might be, respectively, a customer's station and a central office, or the reverse. Or both the source and destination might be central offices. In general, an extremity of a signaling system is required to act as both source and destination with respect to the complementary functions at the other extremity.

The relationship of the signaling system to the transmission media constitutes another important factor in the signaling philosophy. Since the transmission facilities in a telephone system are primarily intended to permit customers to talk to each other, the signaling engineer has the choice of signaling over those facilities within the voice-frequency band or, alternatively, signaling outside of the voice band. If the former

choice is elected, one has the satisfaction of feeling that signaling can be accomplished if talking is possible, but is faced with the problem of signal mutilation and signal imitation due to noise or voice-generated signals. If the out-of-band technique is adopted, the signaling engineer is faced with not only the problem of economically justifying additional signaling channels but also the problem of coordinating voice and signal channels.

In some cases there is little left to choice, and this aspect is accepted as part of the signaling philosophy. For example, the transmission plant involved in the interconnection of switching offices might be so rigidly designed and installed that in-band signaling would be the only economic choice. On the other hand, the choice of out-of-band signaling might be patently simple, such as in the case of DC signaling over wire conductors.

For the most part, the prevailing Bell System philosophy is to employ in-band signaling over nonmetallic facilities and out-of-band (DC) signaling over metallic paths. There are departures from this, and these, along with other factors influencing the application of the over-all signaling philosophy, will be discussed later.

In contemplating telephone signaling systems, it is convenient to make a distinction between the systems employed to signal between the customer and the central office and those used to signal between central offices. This distinction is not always appropriate where step-by-step switching is concerned. However, with modern common control central office switching systems, segregation into these two classifications is quite valid, since the central office usually acts as a buffer between customer stations. In the development of modern signaling systems, it has been desirable also to make this distinction between customer-to-central-office systems and central-office-to-central-office systems in order that an optimum choice of signal encoding, generating and decoding arrangements could be made for each category.

The characteristics of customer-to-central-office systems are influenced by a somewhat different environment than that of interoffice systems. For example, conditions under which signal transmission must be accomplished differ, and fewer information requirements exist in the customer-to-central-office environment. In addition, the need for simple signaling transducers (dials, pushbuttons, bells, etc.) at the customer terminals results in a considerably different approach toward customer signaling systems than toward interoffice systems.

As a result, customer-to-central office telephone signaling systems have tended toward relatively simple arrangements which, with minor varia-

tions, have been standardized for all central office switching systems. Confined to operation over DC loops, except for special cases, these systems employ two-state DC (off-hook, on-hook) signals, rugged DC dial pulse generators and simple AC-operated bells. Party identification and selective signaling for multiparty lines represent signal variations unique to this class of signaling system.

It will, of course, be realized that in defining the scope of the customer-to-central-office signaling systems as above, a great deal has been left unsaid concerning the technical problems and solutions involved in obtaining maximum accuracy and range for the customer signaling devices, while at the same time maintaining low cost. It is interesting to note in passing that the long-term standing of the customer's rotary dial and bell combination is being probed by pushbutton dials and tone ringer explorations. We shall confine the remaining discussion of signaling in this paper to the systems used to communicate between central offices, particularly those of the common control type.

From the foregoing, the essential elements of the signaling philosophy may be summarized as follows:

i. Signaling system objectives are integrated with the prime telephone system technical and economic objectives.

ii. Signaling system requirements are strongly influenced by, and coordinated with, switching and transmission system techniques and requirements.

iii. A large degree of mutual independence is maintained between signaling systems used between customers and central offices and those used between central offices.

In the development of this philosophy a number of principles have emerged. They will be stated without discussion below, but the reader will wish to bear them in mind during the remainder of the discussions and descriptions appearing in this paper.

1. It is desirable to consider a signaling system as a *conveyer* of information, separating out and locating the basic information generators and receivers in other parts of the telephone system to which the signaling system is connected. This permits flexible and interchangeable use of signaling systems with various switching systems.

2. It is not necessary nor desirable to "package" all of the required interoffice signaling functions in one system. On the contrary, separation of functions in such a way that some are associated with the switching system and others associated with an interoffice signaling "link" often permits economies, promotes flexible application of the signaling link, and permits adaptation to different signaling "languages."

3. It is desirable to employ the concept of individual signaling links between switching points, with signal interception and repetition occurring at the switch point. In contrast to end-to-end signaling over built-up connections, the use of individual signal links allows information to be modified or inserted at switch points, and permits the use of different signaling "languages" between switching offices in a built-up connection. The signal transmission problem is also minimized.

4. Signaling links which are alike or symmetrical at each terminal are desirable in order to permit reversible use on two-way circuits.

5. The use of simple two-state signaling to accomplish the control and supervisory signaling functions is desirable where possible to simplify the signaling system and improve its operating margins.

6. "Full duplex" or completely independent two-way handling of trunk control and status signals between offices is desirable for simplicity, speed and accuracy.

7. The use of self-checking or error-detecting codes for the transmission of multibit information, such as the address, has proved desirable from technical and economical standpoints.

8. In-band signaling is desirable where application to voice channels is possible. Special cases of out-of-band signaling, such as DC transmission of trunk control and status signals over available metallic facilities are difficult to improve upon, however. New carrier system techniques, now under development, may influence this picture.

IV. FACTORS INFLUENCING APPLICATION OF SIGNALING PRINCIPLES

4.1 *General*

The function of a signaling system, as noted previously, is to accept source information, convert it to appropriate signals for transmission over telephone lines or trunks and deliver it to its destination. The coordination of this function in a reliable and economical manner with switching and transmission system parameters constitutes the signaling problem. Criteria to be observed in the solution of the problem include speed, accuracy, efficient use of frequency spectrum, provision of adequate signal sensitivity and resistance to signal imitation and mutilation in the presence of noise and other transmission impairments.

The signaling principles outlined in the previous section have been evolved in the course of developing solutions to the signaling problem for various switching and transmission systems as they were conceived for the Bell System. Characteristics of particular switching and trans-

mission systems have in some instances emphasized certain of the principles and in other cases have limited their application. It is interesting to observe that the traditional distinction between "exchange" and "toll" signaling has probably resulted from varying and changing emphasis given to some of the factors next discussed. Parenthetically, it will be noted that this distinction is beginning to disappear, as direct distance dialing (DDD) and carrier-type transmission systems become more prevalent.

Section V of this paper contains descriptions of modern signaling systems as they are employed in the Bell System. The following discussion is intended to bridge the gap between the philosophical concepts discussed up to this point and the practical embodiments of signaling systems.

4.2 *Information Requirements*

Signals destined for customers at the ends of a telephone connection generally are conveyed in the form of audible tones — for example, dial tone, busy tone and ringing tones. These tones also provide information to operators. All other signals are of the type requiring machine recognition or response, and these are primarily of interest in considering the effect of information requirements on signaling systems. The general nature of this type of information is as follows:

(a) information to *control* the seizure, holding or release of an interoffice trunk and the far-end equipment;

(b) information to indicate to the originating end the *status* of equipment at the distant circuit terminal;

(c) information describing the telephone number or telephone *address* of the called customer.

4.2.1 *Control*

Trunk *control* information may be manifested (or registered) by various electrical conditions in a telephone system. For example, the information may be registered by the position of a telephone switchhook, an operated key, a plug inserted in a jack or an operated relay in some part of a switching train (or by a memory cell in electronic systems). In common control systems, the control registration also exists during the addressing period in the device known as a sender.

The nature of the trunk control information is such as to be readily conveyed by a two-state signaling system, and Bell System signaling arrangements are designed to handle the function in this manner. Fur-

ther, the control information is capable of being transmitted continuously following this line of reasoning: let seizure be denoted by a change of state; then hold is a continuance of the state of seizure, with release a change of state in the opposite direction and idle a continuance of the state of release. This method may be contrasted with "spurt" signaling, wherein a change of state is indicated by a uniquely encoded signal of short duration for each state indicated. For simplicity and reliability, most Bell System signaling arrangements have employed the "continuous signaling" mode of operation, although, as shown later, continuous signaling is not applicable to transmission systems where trunk-to-channel concentrations are achieved through speech-interpolating arrangements.

4.2.2 *Status*

One purpose of *status* information is to indicate whether the called customer has answered or not. It is necessary to transmit this information to the originating end of a connection in order to indicate, for charging purposes, the start and end of the conversation time. The source of the information is the called customer's switchhook — either the telephone is off-hook or on-hook (similar information is also generated by the plug-in-jack response of a PBX operator answering a central office call).

As in the case of the control signals, status information indicating the answer condition can be readily conveyed by a two-state signaling system, and signaling systems in the Bell System are designed accordingly. However, in order to convey by two-state systems additional information relative to the status of the called customer's line, such as line or path busy, distinctive signals denoting these conditions are transmitted by coding the on-off intervals. It is interesting to observe in this connection that existing toll signaling systems have been designed so that busy information is conveyed by changes of state in the signaling system. This is done in order to flash an originating operator's supervisory lamp. However, with the trend toward DDD and away from operator-handled traffic, this requirement is being relaxed in new systems since such information can be satisfactorily conveyed to customers by audible tones on direct-dialed toll calls just as it is on direct-dialed exchange calls.

With common-control-type equipment at the terminating end of trunk circuits, additional status information is required to indicate the availability of the "sender" and its readiness to accept pulsing. In this case,

in order that the transmission from the originating to the terminating sender can be coordinated, status signals are returned from the terminating end to provide for "delay dial," "start dial," "stop dial" and "go." These status signals are also two-state in nature, at times being originated in the distant trunk relay equipment and at other times in the distant sender itself.

From the above it will be seen that the information source for both *control* and *status* signals occurs in various parts of a telephone system. The design of telephone switching systems is such that control and status information is usually registered or repeated in trunk equipments associated with each end of an interoffice circuit. The signal function is usually combined with the functions performed by these trunk equipments if a derived interoffice signal link is not required. This is illustrated by interoffice exchange trunks and toll-connecting trunks operating over DC trunk facilities. If, however, the transmission path requires derivation of a signaling link [such as composited (CX) or single-frequency (SF) links to be described later], it is not essential to combine the trunk or other information registers with the signaling link; in fact, it is usually better to separate the signaling link system so that it can be flexibly adapted to the trunk circuits of various switching systems.

4.2.3 *Address*

Let us turn now to the *address* function. The prime source of this information is the originating customer's dial. In nonsenderized step-by-step systems this source is usually repeated by a relay before being presented to the interoffice trunk conductors. Since the address generated by the dial consists of two-state signals, these can be conveyed through later stages of the switching system by any two-state signaling system.

In common control or senderized systems the address information generated by the originating customer's dial is registered on relays in the sender, and this constitutes the information source for subsequent signaling. It is possible to retransmit this source information in various ways, such as by dial pulsing, revertive, multifrequency or other techniques, (concerning which more will be said later). With the exception of dial pulsing, these read-out methods involve conversion to multistate signals. Thus, the effect of these arrangements is to require that ensuing signaling paths be capable of transmitting signals that are multistate in character.

From the above we may conclude that two-state signaling is adequate for all information requirements of nonsenderized systems and, in addition, is adequate for control and status signals of senderized type systems. Practical signaling systems have been designed with this in mind. For the transmission of address information by senderized systems, however, additional signaling capabilities are required. These are discussed in Section 4.4.

4.3 *Switching Systems*

The effect of switching system design on signaling systems is closely related to the effects just discussed. Nonsenderized step-by-step systems need to transmit only simple trunk control signals and trains of dial pulses in the forward* direction, sending status information in the backward direction. While the dial-pulsing function requires careful design and engineering to cope with the problems of pulse distortion and switch limitations, the signaling system as a whole is relatively simple. Common control switching systems increase the effectiveness with which information can be exchanged between switching offices. They also enlarge the opportunity to use improved signaling systems. However, they add considerably to the signaling system requirements.

The increased signaling possibilities of common control systems are obtained through the use of senders, which are digital storage and pulsing devices that can be temporarily associated with trunks and lines during the addressing process. They possess the capability of registering address information and transmitting and receiving this over trunk circuits in various ways. (Receiving-end senders are more appropriately called "registers" in the newer common control systems.) The use of senders in this manner has several effects on signaling systems:

(a) Full duplex signaling is required over the interoffice signal path in order that the sender control signals ("start" and "stop" dial, etc.) described in Section 4.2.1 can be effective.

(b) Full signal regeneration of the address is possible at each switching point.

(c) Higher-speed transmission is made possible.

(d) The addressing system can be separated from the trunk control and status indicating system.

(e) New signaling codes (or languages) can be introduced readily.

The significance of the first four effects noted above is obvious. The last-named effect warrants some further discussion.

* The term "forward" coincides with the direction of traffic over the trunk, the reverse direction being designated as "backward."

4.4 *Signaling "Languages"*

Signaling codes employed by senders in the Bell System for addressing or directing the positioning of switches include the following:

4.4.1 *Dial Pulsing*

Dial pulsing or decimal trains of pulses correspond to those produced by a rotary dial. Senders capable of decoding these pulse trains are used in all common control systems to record the number pulsed by the originating customer. Senders of the same general type, and also those capable of producing dial pulse trains, are required at common control offices when signaling to and from step-by-step systems. For engineering convenience, dial-pulse signals are sometimes used between common control offices. Some dial-pulse senders have been designed to operate at twice the nominal (ten pulses per second) dial speed, and signaling systems are affected accordingly. Otherwise, dial pulse simulation by senders has no effect on signaling systems except as noted in the previous general discussion on sender operation.

4.4.2 *Revertive Pulsing*

Revertive pulsing, the signaling language born of the panel switching system, is so named because of its reverse method of acting. Its principle of operation is broadly as follows: At the terminating end of a trunk a switch composed of a brush having access to a plurality of terminals is set in motion under its own power and is caused to leave a trail of pulses while seeking a terminal. These pulses are received by a sender at the originating end which compares the information they contain with previously stored digital information. When a match is obtained, the sender causes the switch to stop and close its contacts to the terminal at that position.* Information concerning the progress of the call from switch to switch is sometimes required by the sender, and this is transmitted in the same direction as the pulses just mentioned. Thus, the revertive system requires the transmission of three signaling states, two to define the digit pulses and a third for supervision. This has the effect of limiting revertive signaling to circuits specially designed for three-state operation. Up to recently, this has restricted revertive signaling to DC loops, and has prevented its application over derived two-state channels such as com-

* It is of interest to note that the revertive pulsing system represents an early method of accomplishing functions similar to those now performed by modern servo-control systems.

posite signaling, N carrier signaling and single-frequency signaling. However, as described in Section V, the latter has recently been adapted to three-state signaling for revertive pulsing.

Of interest is the fact that revertive pulsing is a substantially higher speed signaling system than dial pulsing. This follows from the fact that the pulses are not required to drive a switch, but are generated by a switch in motion. However, since the sender must tell the switch when to stop, and since this must occur on a terminal corresponding to the revertive pulse count, revertive operation introduces an additional "round trip" time requirement which must be met by the signaling system to keep the method operative. Although revertive is no longer the preferred method of pulsing for new systems, it might outlive the panel switching system, since it has been carried over into crossbar systems, where it presently serves to provide a common language between panel and crossbar and from crossbar to crossbar offices.

4.4.3 *Call Announcer and Call Indicator*

The introduction of dial switching systems into a manual telephone exchange network required that some method be provided to permit a machine office to complete calls to customers still served by switchboard operators. Two methods were devised to accomplish this. In the *call announcer* scheme the address digits were transformed in the machine office to spoken digits, which were then transmitted to the terminating switchboard operator. Another method caused the digital registrations to be transmitted to the terminating switchboard and there displayed as illuminated digits. The display method was called *call indicator*.

The call announcer method, utilizing a photoelectric digit-to-voice encoding machine, was quite ingenious but had limited application and is now practically nonexistent. It will therefore not be discussed further.

The call indicator method of signaling between dial and manual offices was applied both to the step-by-step systems and to the panel system. In the step-by-step system the signaling was done with conventional dial-pulse trains which were decoded and displayed at the switchboard. Since this involved no new signaling techniques, it is of no further interest in this discussion. However, in the panel system the signaling is of interest, since it was accomplished by a new multistate code. (The system is described in detail in Section V.) The *panel call*

indicator (PCI) code has an advantage in speed over dial pulse signaling; however, its multistate characteristics limit it to loop transmission paths capable of accepting multistate signals.

The PCI system has had a significant effect on signaling systems, far beyond its original display purpose. Since PCI senders were designed to transmit both the central office designation as well as the four numerical digits of the customer's number, they were utilized to transmit this information to tandem-switching-type offices when required. Crossbar system senders were also designed to follow suit, both with respect to call indicator operation as well as for operation with tandem offices. Thus, a considerable number of PCI signaling applications have resulted. However, PCI pulsing to manual offices will disappear with manual operation, and its use in signaling to tandem offices is being gradually superseded by multifrequency pulsing as crossbar and panel offices become converted to ten-digit dialing. For this reason, PCI signaling has not been included in the program of adapting exchange-type signaling to AC transmission over carrier systems.

4.4.4 *Multifrequency*

Multifrequency signaling is the most recent Bell System addressing language. Principally used at first as a signaling system for outward calls to the No. 4 toll switching system originating at DSA and CLR boards, and for inward call completion between manual and No. 4 toll switching systems and local crossbar offices, it was rapidly extended to permit distant toll operators to key-pulse calls also to the No. 4 system.

Multifrequency pulsing systems transmit each digit of an address with a single AC spurt. Each digit is composed of a unique combination of two out of a possible five frequencies in the voice band. In addition to the fundamental advantage of making it possible to signal over any voice channel, MF pulsing provides the advantage of being adaptable to operator key pulsing and high speed interoffice signaling. Through the design of its code, the received signals can be checked for parity and rejected if more or less than two frequencies are received for each digit.

Multifrequency pulsing is now standard for intertoll trunks between common control switching systems and for exchange trunks between No. 5 common control switching systems. Systems arranged to use incoming and outgoing multifrequency pulsing presently include the crossbar tandem system, the No. 5 crossbar system, and the No. 4 type

toll system. The No. 1 crossbar system is arranged for completion of terminating calls from offices equipped to send multifrequency pulsing. The older systems such as panel and step-by-step still depend upon their original methods of pulsing except that both panel and No. 1 crossbar systems are now capable of seven- and ten-digit multifrequency pulsing to tandem switching systems arranged for centralized AMA.

Multifrequency signaling, as indicated above, has had a considerable impact on the signaling arrangements used for switching systems. Its use is continuing to grow and new applications continue to be found. Systems based on the multifrequency pulsing objectives are now being considered for customer-to-central-office pushbutton signaling. In this connection, the use of a different coding system is being considered, but loss of compatibility may not be objectionable since, as noted earlier, the central office serves as a buffer between customer signaling systems and interoffice signaling systems.

4.5 *Other Switching System Effects on Signaling*

Returning now to the general matter of switching systems effects on signaling, we observe at least one other broad effect which is worth dwelling upon, i.e., one-way versus two-way operation of interoffice trunk circuits.

From the standpoint of traffic, it has been found satisfactory to operate with one-way trunk groups in exchange switching. In this method of operation, traffic always originates in the same direction over a trunk group. Trunk and signal circuits designed for one-way operation tend to be simpler than for two-way operation since they can be designed to fit only the signaling requirements that are unique to each end of the circuit.

Two-way operation has been the usual practice for long-haul or toll-type circuits in order that these costlier circuits can be used more efficiently. In practice, the additional signaling and trunk-switching complexity has been absorbed in the trunk equipment at each terminal, and the signaling arrangements have been kept simple by designing them for symmetrical operation at each terminal. With symmetrical operation, each of the input-output functions is applied, conveyed and delivered in an identical manner for each direction of operation. Since for the longer circuits the control and status functions are always conveyed over a derived signal link, such as CX or SF, and since these signaling functions have been reduced to two-state conditions in either direction it has been possible to design the signal links for symmetry without penalty.

4.6 *Effect of Special Signal Requirements on Signaling Systems*

The reader will note that these effects might have been considered in preceding sections. In electing to discuss them separately, we hope that there will perhaps be a little gain in clarity.

For the purpose of this discussion "special signal requirements" are considered to include such signaling functions as are not directly performed in conveying the basic information involved in trunk *control*, far-end *status* and called telephone *address*. Some special signals presently in use which are of interest are:

(a) originating end re-ring (ring forward);
(b) auxiliary charge functions;
(c) interoffice coin (box) control;
(d) calling party identification.

Of these, only the first directly affects the basic interoffice systems used for control and status. The remaining categories are conveyed by separate signaling systems which require only that they be mutually noninterfering.

4.6.1 *The Re-ring Function*

This is required only when an originating operator desires to attract the attention of a terminating operator after a connection has been established. It will be noted that this is never required on a direct customer-dialed call, either exchange or toll. And it is required on an operator dialed call only when the call cannot be directly dialed to the called customer. In this case an "inward operator" is called in at the terminating end to handle the call. The re-ring signal is a carry-over from manual ringdown practice, and will probably disappear when full DDD is in effect.*

The re-ring heritage goes back to the time when it was conveyed by a ringing signal which was the same whether used for connect, disconnect or re-ring. In carrying this function over to automatic signaling systems, it was integrated with the two-state systems which, it has been observed, are otherwise satisfactory for interoffice control and status requirements. This was accomplished by making use of time to derive a third state. Thus, to the basic connect and disconnect states, a third re-ring state, consisting of a short pulse of the disconnect state, was added. The effect of this is that signaling and switching systems must be designed with a proper regard for time, the signal system

* The re-ring function will be retained, however, for special operating assistance on dialed international calls, as discussed in Section VI.

being required to accurately sense and convey re-ring impulses, and the switching system being required to distinguish between short re-ring pulses and long disconnect signals. Obviously this requires close coordination between signaling system and switching system design, and it has been the basis of some interesting technical problems during the introductory coordination of new signaling systems with old switching systems.

4.6.2 *Remaining Special Signals*

We shall discuss these only briefly. Signals for charging purposes normally are conveyed by the called party answer and disconnect signals. These are interpreted in the originating switching office and result in the operation of registers or in records being made that correspond to the time of answer and disconnect on automatic message accounting tape. The primary effect on signaling systems of these charging systems is to make it quite essential that an answer signal be properly distinguished from a flashing signal, and this is done by appropriate timing in the trunk equipment at the originating switching office.

There exists in a few of the large metropolitan areas a system known as "remote control zone registration." In this, common equipment at a tandem switching point is used to generate a charge signal, depending on the destination of the call and the elapsed time of conversation. This information, consisting of a measured number of pulses, is transmitted from the tandem office back to the originating office where the customer's message register is operated. It is essential, of course, that this method be impervious to signal imitation by other signaling or switching functions. Hence, as applied in DC exchange signaling systems, the remote control signals consist of relatively high voltage pulses which are not readily imitated.

4.6.3 *Other Applications*

Some additional applications of the multifrequency system recently coming into use relate to the third and fourth categories of special signals noted above. In case (c), signal pulses similar to the digits in the multifrequency code are used to send "coin collect" or "coin return" signals from a master office to a satellite office not equipped with local coin supervisory equipment. In case (d), the identity of a line originating a call is transmitted using the multifrequency code from an originating switching office to an office equipped with centralized AMA recording equipment. These systems are chiefly of interest in this

discussion in illustrating the basic utility and versatility of the multifrequency method of signaling. The primary impact of both of these systems on the interoffice signaling system is that they require careful coordination therewith. It is obvious that this is a fundamental rule that must be observed in considering the application of any new signal or signaling system to the telephone plant.

4.7 *Transmission Media*

The influence of the transmission media on the development of signaling systems has been at least as significant as that exerted by the switching considerations just discussed. The most significant impact, of course, is in the transition from simple metallic voice circuits, served by DC signaling, to repeatered or carrier-type facilities requiring AC signal transmission.

The DC signaling design problem has been somewhat similar to that encountered in the design of DC telegraph facilities. Signaling arrangements using both neutral and polar transmission have been employed in the design of trunk circuits operating over metallic facilities, with the choice being based on range requirements and economy. Where distances between central offices indicated the need for a higher degree of sophistication, other telegraph techniques were introduced in the form of simplexed and composited signaling arrangements.

As dialing distances increased and the transmission path required the use of voice-frequency repeaters or carrier systems, it became necessary to either transmit the signaling information over paralleling DC telegraph facilities or by AC signals over the same path used for voice. The extensive use of carrier systems, in particular, dictated that AC signaling be employed.

Signaling systems using low frequencies, such as those used for early ringdown signaling systems, cannot be applied on circuits with composited telegraph. Also, low-frequency signaling systems are relatively slow in operation, due to lack of sensitivity and the need to protect the circuits from speech imitated signals. Even the 1000-cycle ringdown signaling arrangements, using 20 cycles as a modulating frequency, are too slow to be used for dial pulsing. Accordingly, AC signaling systems for dial operation required a new approach. In considering such systems the question is encountered whether to use frequencies within or outside of the voice band. Some of the advantages and disadvantages of the alternative schemes are reviewed below.

The obvious advantages of out-of-band signaling are, of course, free-

dom from the effects of speech currents, compandors and echo suppressors. In general, out-of-band signaling also allows relatively simple terminal equipment and has the additional advantage of permitting signaling to take place during the talking interval. The necessity of providing additional bandwidth, filters to provide the signaling slot, and the possibility of switching to a trunk over which it is possible to signal but impossible to talk due to a trouble in the separate speech path, are disadvantages.

In-band signaling systems have the advantage of not requiring additional bandwidth. In addition, since the signaling currents utilize the same path as the speech currents, the amplification provided by voice amplifiers also renews the strength of the signaling currents. With in-band signaling it is possible to quickly substitute another voice channel in case of trouble, since the signaling is carried along with the speech circuit.

In-band signaling systems are advantageous when telephone facilities within a trunk are connected in tandem. Under these conditions signaling equipment need only be furnished at the terminals of the trunk and not at the intermediate point within the trunk. This is economically significant in the Bell System toll plant, where an average of 1.4 transmission links comprise a trunk. In addition to the economic advantage, there is no additional signal distortion introduced through equipment at intermediate points.

In-band signaling systems tend to be more complex because of the need for protection against speech-generated signals, but the Bell System type of in-band signaling provides protection adequately and economically by signal-to-guard arrangements in the signaling receiver. The guard uses the energy present in the frequency spectra outside of that assigned to the signaling frequency itself. Such guard energy, when combined with the energy present at the signaling frequency, determines whether the receiver is to operate or not. The ratio of signal-to-guard voltages can be adjusted at the time of design of the receiver, and this, together with timing, provides protection against speech operation.

In choosing the frequency to be used for supervisory signaling it is desirable to select as high a frequency as possible in order to reduce the occurrence of speech-imitated signals, since speech energies at the higher frequencies are, of course, a good deal less than those present in the lower ranges. In addition, it is desirable to select a frequency that will not interfere or be interfered with by signaling systems using other tones which may be present at the same time such as multifrequency pulsing arrangements. The power level of the signaling frequency should also be low enough to avoid overloading the intermediate amplifiers, but it must

be of sufficient level to provide satisfactory operation of the receiver in the face of the usual types of noise present on telephone trunks.

The earliest in-band signaling system utilized a frequency of 1600 cycles. This frequency was dictated by the use of "emergency bank" carrier channels which were used during World War II to provide additional circuits. Trunks which included such filters had an upper frequency cutoff of about 1750 cycles, and hence a signaling frequency of 1600 cycles was selected. With the close of the war and the removal of the emergency equipment it was possible to raise the signaling frequency to 2600 cycles. At this higher frequency the problem of protection against speech operation was made considerably easier and permitted the use of smaller components as well as less complex circuitry. Present standard voice-frequency signaling systems employ tones of 1600 to 2600 cycles and use all frequencies, except the signaling frequency, to guard against false operation of the signaling receiver on speech currents.

Other AC signaling systems such as the multifrequency pulsing system must also be designed with the transmission media in mind. Falling in the band between 700 and 1700 cycles, they are just within the bandwidth provided by the emergency facilities and well within the voice band of standard voice circuits.

The most recent extension of AC signaling has been in connection with the exchange plant. With the continued trunk growth and increasing costs of voice-frequency cable facilities, the use of short-haul carrier systems appeared to offer economic advantages in the exchange plant over physical facilities. With the development of short-haul carrier systems arose the need for inexpensive signaling arrangements to handle the interoffice trunk signals.

Initially, two-state out-of-band signaling was employed using a signaling slot outside but adjacent to the voice band, and a system using 3700 cycles was designed for types N, O and ON carrier systems. While the signaling devices for this system were less expensive than in-band signal circuits, the system lacked the in-band advantages discussed earlier; also, the loss of circuit patching flexibility limited their usefulness somewhat. For these reasons, and in anticipation of the large-scale use of carrier in exchange areas, new transistorized in-band signaling arrangments were designed. The new in-band systems not only accommodated the two-state signaling required for control, status and dial pulse address signals, but also were adequate for the three-state revertive pulsing signals required by panel and some crossbar offices. In addition, the new designs were adapted to operation with a variety of switching system trunk circuits.

4.8 *Compatibility*

During the evolution of the Bell System plant it has been almost axiomatic that new systems should work with existing systems. This follows from the fact that it is difficult to justify economically the abandonment of existing plant in favor of permitting a new system to function in an environment all its own. Thus arises a problem in compatibility.

Prior to the widespread application of dial systems the compatibility problem was not severe, since intersystem signaling requirements for ringdown operation were relatively simple and dial system isolation was possible to a considerable degree. However, as the total number of dial central offices increased and, as the varieties increased, the compatibility problem became more important, particularly where direct dialing exchange areas were increased in size and scope. Finally, the introduction of nationwide dialing concepts caused the compatibility problem to become full blown.

Where the evolutionary developments in switching systems required that these systems communicate with each other, it was necessary to accomplish this over transmission facilities which also were undergoing evolutionary changes. It is obvious therefore that systems used for intersystem signaling have been faced with a sizable compatibility problem. In the solution to this problem it has been necessary also to make sure that signaling systems were properly coordinated with each other in order to avoid interference. Finally, the solution to the compatibility problem has required that the design of signaling systems adequately provide for flexible plant administration methods.

Up to about the beginning of the last decade, dial switching systems used for exchange service were interconnected over metallic paths and consequently employed DC signaling methods. Since, for operation over metallic facilities, the "control" and "status" functions are usually combined with the switching system trunk equipment, and since these equipments usually contained features unique to the switching system, there appeared to be little need for a systematic approach toward providing compatible signal circuits for interchangeable use between different switching systems. Consequently there resulted a large number of trunk circuit designs, many of which were uniquely tailored to a particular switching and signaling plan.

The extensive interconnection of different kinds of dial switching systems and the requirement for extending signaling ranges occurred at about the same time. Initially the extension in signaling ranges was obtained through the use of telegraph-type derived signal links which were

obtained by compositing the voice circuits. These signal links (CX) presented the opportunity for providing signaling compatibility. Thus resulted the standardization of input and output signals at the terminals of derived signal links. Output and input signaling leads, called somewhat arbitrarily E and M leads respectively, became standard terminations for CX signaling circuits and provided the capability of always delivering and accepting uniform signal conditions. This input-output standardization permitted CX signal links to be connected to corresponding E and M lead trunk circuits of any system. These standard terminations were carried over to the signal links later developed for use over nonmetallic facilities such as the various single-frequency signaling systems and out-of-band signaling systems.

As noted previously, the use of senders in common control offices has led to the use of a number of signaling languages for the address function — revertive, PCI and multifrequency. In order for systems to communicate with each other they have been equipped in varying degrees with registers and senders to accept these signaling languages. Table I shows the extent to which modern switching systems are able to address each other using the signaling languages. It will be noted that some systems have the ability to communicate in several ways with a particular connected system. The selection of a particular method is left to engineering choice, taking into account the total interconnecting requirements of the exchange or toll office under consideration.

The Bell System philosophy of repeating signaling information at each switching point has assisted in obtaining compatibility between various signaling systems, since at each switching point it is possible to convert the input control, status and address information to meet the requirements of the switching point beyond.

The problem of maintaining compatibility with transmission systems has involved conversion of interoffice signaling systems from DC to AC methods, where carrier transmission facilities were substituted for metallic facilities. With in-band signal system operation, proper coordination in the design of transmission and signaling systems has assured that satisfactory signal transmission will occur over any voice transmission system, and the standardization of input and output signaling leads for CX, SF and out-of-band signal links has made it possible to flexibly interconnect switching systems over DC or AC transmission facilities.

It is of interest to note the degree to which AC signaling systems must, of themselves, observe the rules of compatibility. Fig. 1 indicates the manner in which the voice-frequency spectrum is now utilized for the transmission of AC signals for telephone and other Bell System services.

TABLE I — ADDRESSING SYSTEMS BETWEEN OFFICES

Originating Office	Terminating Office									
	Step-by-Step	Step-by-Step Tandem	No. 4 Toll Crossbar	Crossbar Tandem	No. 1 Crossbar	No. 5 Crossbar	Panel	Office Selector Tandem	Panel Sender Tandem	Local Manual
Step-by-Step	DP	DP	DP	DP	DP	DP	X	X	X	V CI
Step-by-Step Tandem	DP	DP	DP	DP	DP	DP	X	X	X	V CI
Local Manual	DP	DP	DP MF	DP MF	DP MF	DP MF	DP	X	V	V
Dial System A	DP	DP	DP MF	DP MF RP	DP MF RP	DP MF RP	DP RP	RP	V PCI	V PCI
Manual Toll	DP	DP	DP MF	DP MF RP	MF RP	DP MF	RP	X	V	V
Panel Sender Tandem	DP	DP	X	X	DP RP	DP RP	RP	X	X	PCI
Crossbar Tandem	DP	DP	DP MF	MF RP	MF RP	DP MF RP	RP	X	X	V PCI
No. 4 Toll Crossbar	DP	DP	MF	DP MF	MF RP	DP MF RP	RP	X	X	V PCI
No. 5 Crossbar	DP	DP	MF	DP MF PCI	MF RP	DP MF RP	RP	X	PCI	V PCI
Panel	X	X	MF	MF PCI RP	MF RP	MF RP	RP	RP	PCI	PCI
No. 1 Crossbar	X	X	MF	MF PCI RP	MF RP	MF RP	RP	RP	PCI	PCI
Office Selector Tandem	X	X	X	X	RP	RP	RP	X	X	PCI

DP = Dial pulsing; RP = Revertive pulsing; MF = Multifrequency pulsing; PCI = Panel call indicator; CI = step-by-step call indicator; V = Verbal; X = No connection.

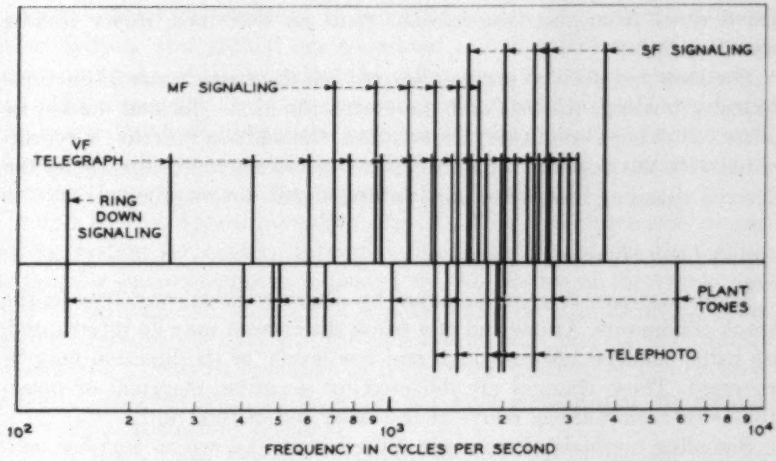


Fig. 1 — Various signaling and plant tones.

It is obvious that where any two or more of these signals are present at the same time it is essential that they be mutually noninterfering.

V. MODERN INTEROFFICE SIGNALING SYSTEMS

5.1 *General*

Up to this point we have discussed the nature and effect of signaling principles on the evolution of the signaling art. An idea of the methods of implementing these principles in practical signaling systems will be obtained from the following general descriptions. These are confined to typical systems and are divided broadly into two categories, DC signaling and AC signaling. In each category the performance of the trunk control, status and addressing functions will be observed, and also the application in the exchange and intertoll plant.

5.2 *DC Systems*

In the exchange plant where short distances are frequently encountered DC signaling systems continue to be used extensively. They are used also for some short-haul toll systems. The systems are of two general types, the simplest occurring where a DC loop path is available. These are grouped below as *loop signaling systems*. Where a clear loop path is not feasible or where extended DC ranges are desired, signal links

are derived from the transmission path as described under *derived signaling links*.

The loop and derived systems accomplish the trunk control functions (seizure, holding, release) and the status functions (far-end answer or other switchhook condition). In addition, the address function is coordinated with the loop signaling arrangements and conveyed directly by the derived signaling links when dial pulsing signals are required.

5.2.1 *Loop Signaling Systems*

These systems, in general, signal by altering the current flow in the trunk conductors. At one end of a trunk the current may be interrupted, its value changed between high and low levels, or its direction may be reversed. These changes are detected by sensitive, marginal or polar-type electromechanical relays at the other end of the trunk.

Signaling methods of this type have become known as *high-low, wet-dry, reverse-battery* and *battery-ground*. The older high-low and wet-dry methods are being rapidly supplanted by reverse-battery and battery-ground techniques.

5.2.1.1 *High-Low Signaling* (*Fig. 2*). This method of signaling is accomplished by marginal current changes. A seizure signal is originated by applying battery and ground to the trunk in series with a marginal relay C. At the distant end, relay L operates and a call indication is given to the called office. When the called customer answers, the S relay operates to short-circuit the high resistance winding of the L relay, causing sufficient current to flow to operate the C relay to indicate the answer to the calling end of the trunk.

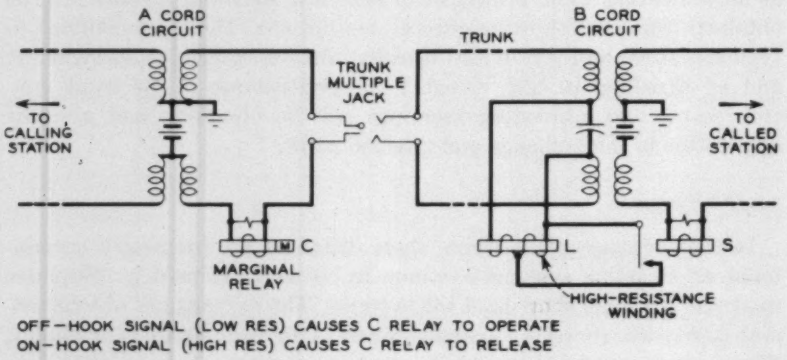


Fig. 2 — High-low signaling.

5.2.1.2 *Wet-Dry Signaling.* In this type of signaling a trunk is "wet" when battery and ground are connected to the called end of a trunk and a DC bridge is connected at the calling end to initiate a seizure signal. The customer answer removes the battery and ground, and hence the trunk is "dry" during the busy or off-hook condition and "wet" during the idle or on-hook interval. A release signal is registered by opening the DC bridge at the calling end of the trunk.

5.2.1.3 *Reverse-Battery Signaling (Fig. 3).* This is the preferred DC signaling system for modern interoffice exchange trunks. At the calling terminal a seizure signal is indicated by the closure of the trunk conductors through the windings of a polar relay CS. The resulting trunk current operates the A relay at the called terminal. Upon called customer answer the T relay is operated to reverse the polarity of the battery, which in turn operates the CS relay. A release is indicated by opening the trunk at the calling terminal.

5.2.1.4 *Battery and Ground Signaling.* A variation of reverse battery signaling is obtained by supplying battery and ground at each end of the trunk series aiding. This effectively doubles the value of the trunk current and permits increased operating range.

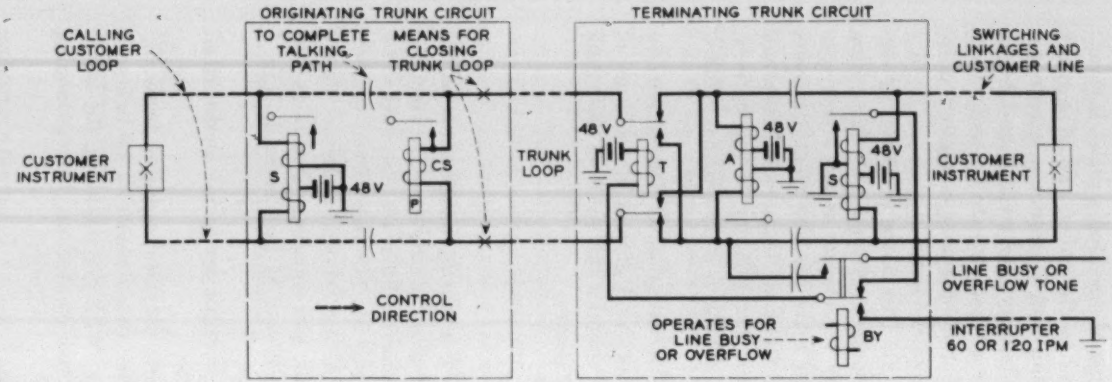
5.2.2 *Derived Signaling Links*

Derived signaling links are used for the longer exchange plant trunks and for short-haul intertoll trunks. Signaling connections between the trunk relay circuits and the derived signaling links are obtained via a uniform system of leads designated E and M. For this reason systems of this type are frequently known as E and M lead systems. Fig. 4 shows the interconnection between trunk and signaling circuits as well as the potentials and directions of the E and M lead controls. Fig. 5 shows further detail.

Several types of derived DC signaling circuits are in use as described below.

5.2.2.1 *Simplex (SX) Signaling (Fig. 6).* Simplex signaling feeds signaling currents through center taps of line transformers to the balanced path furnished by the trunk conductors. With SX signaling, the trunk resistance is halved by paralleling the two conductors, thus extending the range as compared to loop signaling.

5.2.2.2 *Composite (CX) Signaling (Figs. 7 and 8).* Composite signaling consists essentially of a high-pass-low-pass filter arrangement which separates the DC and low frequency signaling currents from the voice frequency signals, the separation point being at about 100 cycles per



LINE CONDITIONS								
BEFORE CALLED CUSTOMER ANSWERS	OFF-HOOK	S OPERATED	CS NORMAL	LOOP CLOSED	LOOP NORMAL	A OPERATED	S NORMAL	ON-HOOK
CALLED CUSTOMER ANSWERS	OFF-HOOK	S OPERATED	CS OPERATED	LOOP CLOSED	LOOP REVERSED	A OPERATED	S OPERATED	OFF-HOOK
CALLING CUSTOMER HANGS UP FIRST	ON-HOOK	S RELEASED	CS RELEASED	LOOP OPEN	LOOP REVERSED	A RELEASED	S OPERATED	OFF-HOOK
	ON-HOOK	S RELEASED	CS RELEASED	LOOP OPEN	LOOP NORMAL	A RELEASED	S RELEASED	ON-HOOK
OR								
CALLED CUSTOMER HANGS UP FIRST	OFF-HOOK	S OPERATED	CS RELEASED	LOOP CLOSED	LOOP NORMAL	A OPERATED	S RELEASED	ON-HOOK
	ON-HOOK	S RELEASED	CS RELEASED	LOOP OPEN	LOOP NORMAL	A RELEASED	S RELEASED	ON-HOOK

THE CS RELAY IS POLAR IN OPERATION BUT ALSO HAS MECHANICAL BIAS TO CAUSE RELEASE ON OPEN CIRCUIT

Fig. 3 — Reverse battery signaling.

SIGNAL A TO B	SIGNAL B TO A	CONDITION AT A		CONDITION AT B	
		M LEAD	E LEAD	M LEAD	E LEAD
ON-HOOK	ON-HOOK	GROUND	OPEN	GROUND	OPEN
OFF-HOOK	ON-HOOK	BATTERY	OPEN	GROUND	GROUND
ON-HOOK	OFF-HOOK	GROUND	GROUND	BATTERY	OPEN
OFF-HOOK	OFF-HOOK	BATTERY	GROUND	BATTERY	GROUND

Fig. 4 — E and M lead conditions.

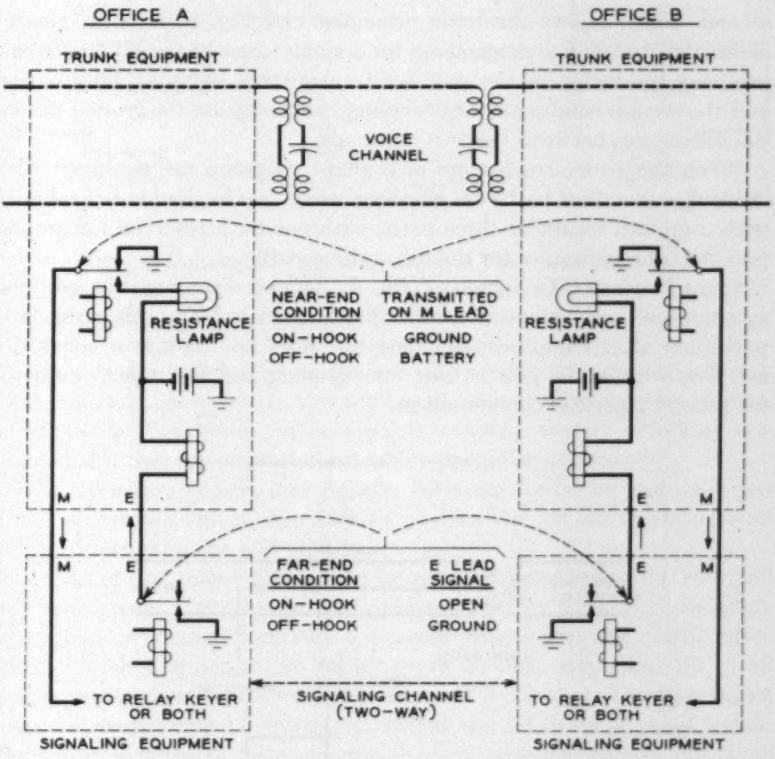


Fig. 5 — Signals between trunk and signaling equipment in E and M lead signaling systems.

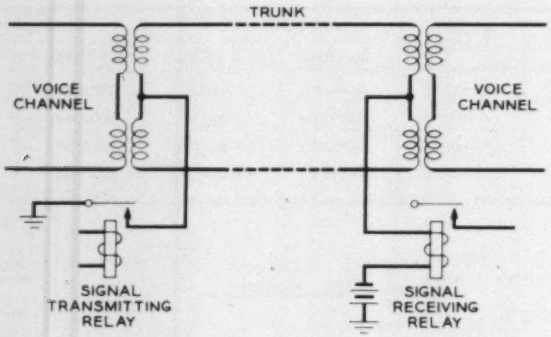


Fig. 6 — Principle of simplex signaling.

second. Fig. 7 shows the basic principles and Fig. 8 shows in greater detail the signaling arrangements for a single voice channel. The circuit uses one conductor of the pair for transmitting signaling information and the second conductor for providing compensation for ground potential differences between the two terminals.

When the trunk conductors of a phantom group are equipped with composite sets, four paths are obtained which can be used independently with a ground return or three paths with one path reserved for ground potential compensation for the other three paths.

5.2.2.3 *Duplex (DX) Signaling (Fig. 9).* Duplex signaling is based upon symmetrical and balanced circuit that is identical at both ends. It is patterned after composite signaling but does not require a composite set. One wire of the pair is used for signaling and the other conductor for ground potential compensation.

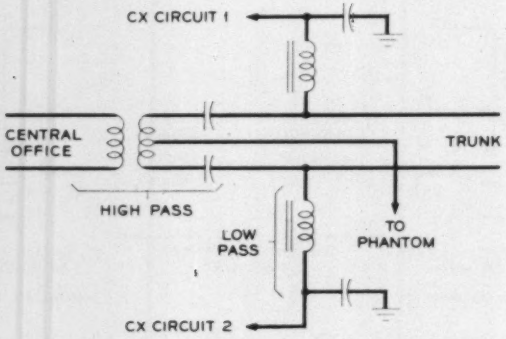


Fig. 7 — CX circuit for one end of a trunk pair.

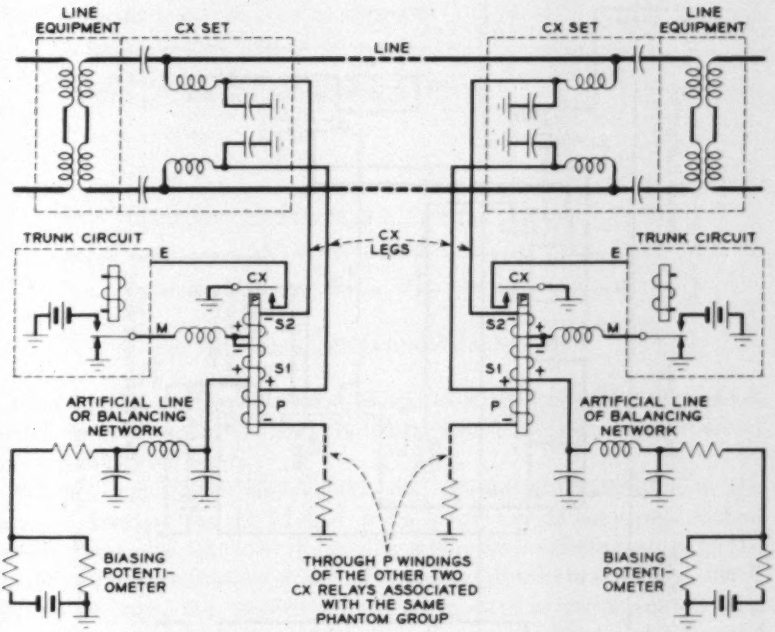


Fig. 8 — Composite signaling circuit for one-voice channel.

5.2.3 *Addressing by DC Methods*

5.2.3.1 *Dial Pulsing.* The customer's dial actuates a cam which generates open and close signals, the number of open or "break" signals being equal to the digit being transmitted. In common control switching systems dial pulses are generated by relay-type pulse generators.

Fig. 10 shows typical dial signals. Between digits the loop is closed for a somewhat longer time, which enables the switch in the central office to recognize the end of a digit.

Trains of dial pulses are defined in terms of pulse repetition rate and per cent break. (Per cent break is the ratio of the duration of a single open loop or "break" interval to the sum of the open and closed intervals.) Nominal speed values for all types of dials vary from 7.5 to 12 pulses per second; breaks vary from 58 to 67.5 per cent. The latest dials vary between 9 and 11 pulses per second and 60 to 64 per cent break. Some special operator dials are capable of 20 pulses per second, but these are only used with compatible central office equipment.

With loop signaling, dial pulses actuate a receiving relay, such as the

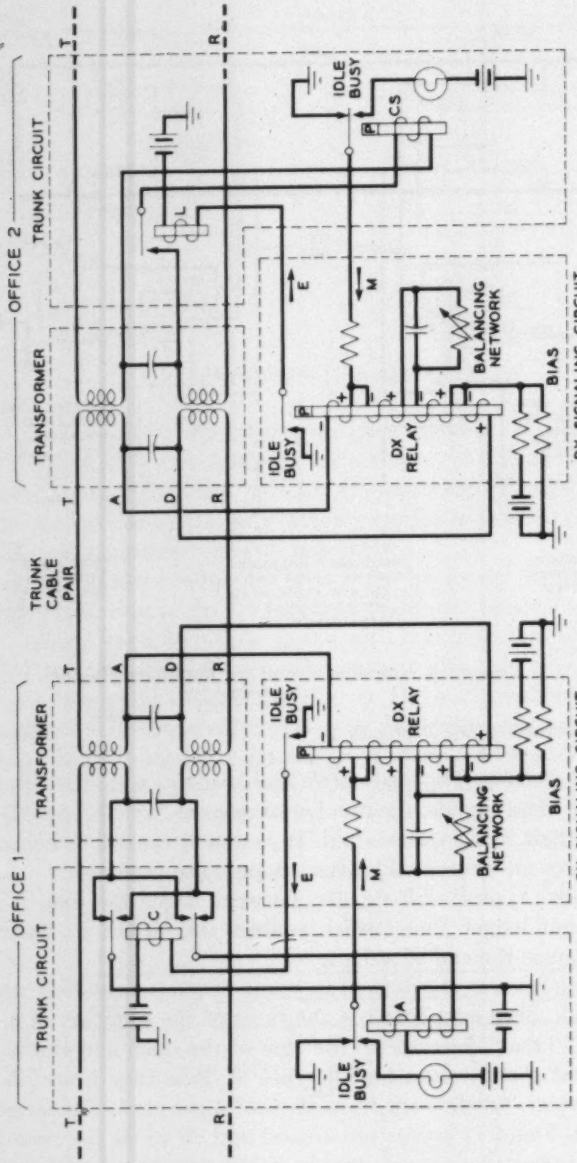


Fig. 9 — Duplex (DX) signaling.

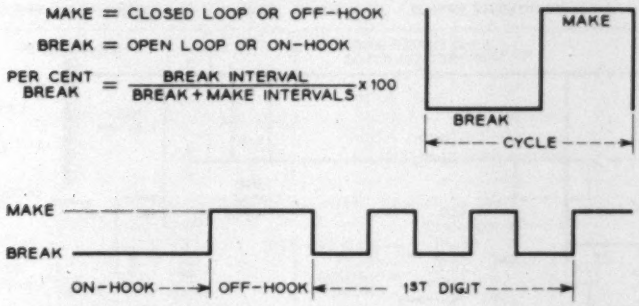


Fig. 10 — Signaling conditions during dialing.

s relay in Fig. 3. The operation of this relay either directly causes a switch to step or, in common control systems, the pulses are counted and registered by relays.

When using derived signal links, dial pulsing is transmitted in the forward direction via the M lead in the same way as the trunk control signals. Pulses are sent over the facility and result in similar pulses on the E lead of the terminating circuit. Through appropriate circuitry, the E lead pulses cause the terminating switch to step, as in the case of loop signaling circuits or, alternatively, the pulses are counted and registered by relays in common control systems.

The design of many of the E and M lead signaling circuits includes pulse correcting circuits to maintain the pulses within the operating limits of the associated central office equipment.

5.2.3.2 *Revertive Pulse Signaling.* As used in the panel system, Fig. 11, a "start" pulse from the originating office causes a terminating selector to be started and driven by its own power over a bank of terminals. The selector signals its position at each step by a pulse which is generated by a grounded brush passing over a commutator associated with the switch mechanism. A counter at the originating end of the trunk counts these pulses and initiates a signal to stop the switch when the desired position is reached.

When this is used in crossbar systems, a relay in the terminating office sender produces a train of revertive pulses upon receipt of a start signal from the originating office. The pulses are counted in both the originating and terminating offices. When a stop signal is received from the originating office, the count in the terminating office is transferred to a register, and the counter is released for use in the next digit.

Revertive pulsing is faster than dial pulsing, but requires signaling

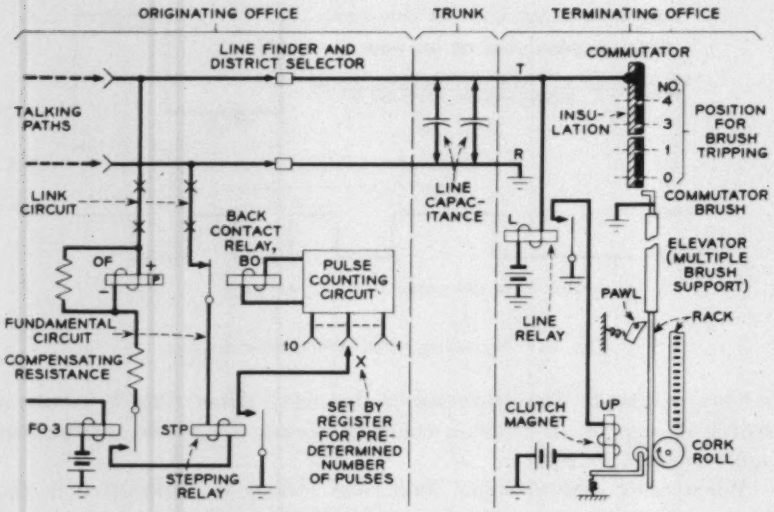


Fig. 11 — Revertive pulsing between panel offices.

circuits capable of operating up to 22 pulses per second from crossbar senders and up to 32 pulses per second from panel selectors.

The signals transmitted over the trunk conductors are shown in Fig. 12. Revertive signaling requires transmission of three states since, in addition to the two states required for the terminal count, a third state is transmitted after all selections have been pulsed. (This is referred to as "incoming advance" and signifies that the switching equipment can cut the trunk through to the talking paths at each end.) The three states are transmitted by two polarities and a zero current condition.

5.2.3.3 *Panel Call Indicator Signaling.* The PCI code is essentially a binary system employing four bits per decimal digit. The two states of

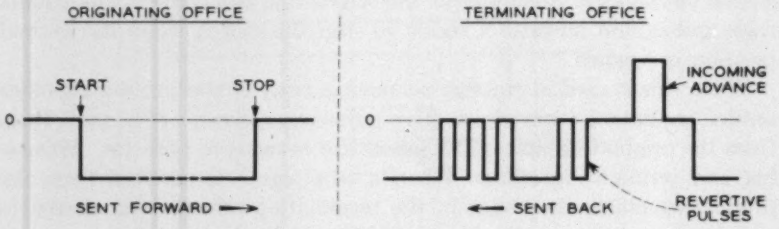


Fig. 12 — Revertive pulsing.

TABLE II — PCI SIGNAL CONDITIONS

Symbols	Loop Conditions
—	open loop
p	light positive current
n	light negative current
N	heavy negative current

(Polarities are ring with respect to tip.)

Basic PCI Cycle				
Time Intervals	A	B	C	D
Normal Pattern	—	n	—	n
Permissible Modifications	p	N	p	N

the first and third bits of each digit are defined by open and positive pulses with the states of the second and fourth bits being defined by light negative and heavy negative pulses. These variations are shown in Table II, and the complete code is given in Table III. The alternate negative bits of the code are used to synchronize the receiving with the sending end and to advance the register to successive digits. Four decimal digits, sent consecutively with no pause in between, require a total transmission time of about one second. The signal characteristic as it appears on the trunk conductors is shown in Fig. 13. A heavy positive pulse is transmitted to indicate end of pulsing. This, together with the four states required to define alternate bits of the code, results in a five-state signaling system.

As shown in Fig. 14, PCI pulses transmitted over the trunk are detected by means of polar and marginal relays at the receiving end.

TABLE III — PCI CODES

Digit	Hundreds Tens and Units				Thousands			
	A (1)	B (2)	C (4)	D (5)	A (2)	B (4)	C (8)	D (1)
0	—	n	—	n	—	n	—	n
1	p	n	—	n	—	n	—	N
2	—	N	—	n	p	n	—	n
3	p	N	—	n	p	n	—	N
4	—	n	p	n	—	N	—	n
5	—	n	—	N	—	N	—	N
6	p	n	—	N	p	N	—	n
7	—	N	—	N	p	N	—	N
8	p	N	—	N	—	n	p	n
9	—	n	p	N	—	n	p	N

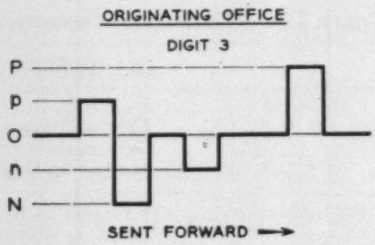
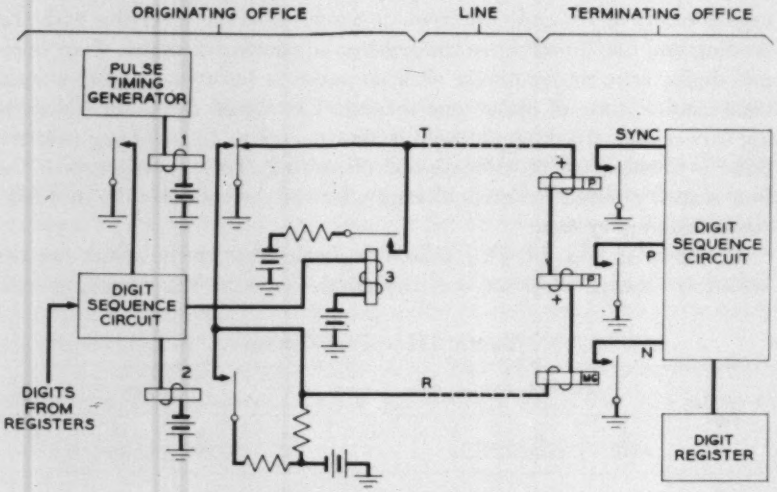


Fig. 13 — PCI pulsing.

The output of the detectors is registered on relays or switches in the terminating office. When the PCI transmission is to a manual office the registrations are displayed on a field of lamps in view of the operator. The display consists of ten numbered lamps for each digit, one of which is lighted to indicate the value of the digit.



RELAY CONDITION			LINE CONDITION
1	2	3	
+		−	−
−	−		n
+		+	p
−	+		N

Fig. 14 — Basic PCI pulsing system.

5.3 *AC Systems*

5.3.1 *General*

Alternating-current signaling systems have been designed to convey the basic trunk control, status and address functions required by switching systems. They are used over toll and exchange trunks where DC signaling is not feasible or economical, such as long-haul circuits and short-haul circuits equipped with carrier. Two-state AC signaling can handle trunk control, status and addressing where the latter is coded by dial pulsing. Three-state AC signaling has been designed to handle the revertive method of addressing, including the trunk control and status signals. Multistate AC signaling, in the form of multi-frequency pulses, conveys the address function only. It is coordinated with two-state trunk control and status signaling, either AC or DC, as required.

In the two-state AC category, systems using both in-band and out-of-band signaling frequencies are in use. In-band systems are frequently referred to as *voice-frequency signaling*. They will be discussed first.

Voice-frequency signaling has a wide variety of applications. Frequencies in the voice band from about 500 cps to about 2,600 cps are used and signaling equipment is required only at the terminals of a long transmission path.

Voice-frequency signals are usually of the same order of amplitude as voice currents so as not to overload voice amplifiers or cause crosstalk in adjacent channels. Consequently, they cannot be detected by present-day electro-mechanical devices but must be detected by electron tube or transistor circuits.

One of the chief problems in voice-frequency signaling is prevention of mutual interference between voice and signals. Voice-frequency signals are audible and consequently signaling must not take place during the time the channel is used for conversation. Signal receiving equipment, however, must remain on a channel during conversation to be ready to respond to incoming signals and it may thus be subject to false operation from voice sounds which resemble the tones used for signaling. Protection against voice interference is accomplished in a number of ways:

1. Signal tones of a character not likely to occur in normal speech are chosen. In so-called 1000-cycle signaling, the frequency is modulated at a 20-cps rate. In multifrequency systems, several frequencies are combined and individual receivers set up to respond to each frequency.

2. Time delay is often used to prevent false operation due to voice. The signal receiver is made slow to respond so that normal noise and voice currents which would otherwise actuate the receiver are ignored.

3. Voice-frequency energy, other than the signaling frequency, is detected and used to inhibit the operation of the signaling receiver.

First single-frequency (SF) circuits used 1600 or 2000 cycles for signaling frequencies, while the latest circuits use 2400 or 2600 cycles. Since these frequencies are in the voice band they can be passed over the line facility with the same ease as voice currents.

The SF signaling system is designed to pass the necessary signals for telephone trunks over voice frequency transmission line facilities without impairing the normal use of these facilities for speech. This system accepts and delivers DC signals from the switching trunk equipment. The DC signals are transformed to AC forms on the line side, and vice versa. The same frequency is employed in both directions on four-wire line facilities, since these have separate transmission paths between terminals. On two-wire facilities, different frequencies are used in the two directions. One signal frequency is applied or removed at each end of a line facility to operate or release a relay at the far end. In this way, two alternate signal conditions are provided in both directions of transmission. Normally, speech and signal frequency are not on the line facility at the same time.

The SF signaling circuit is inserted in series with separate transmitting and receiving branches, that is, the four-wire transmission points at the terminals of line facilities. The office side of the transmitting branch may be opened momentarily for critical signaling conditions, but the receiving branch is provided with continuous, one-way transmission. These features protect the signaling circuits from noise and tones originating on the office sides of each terminal.

Relatively high signal power of short duration is used for effective operation in the presence of greater than normal line noise. The guard-channel principle is employed to avoid false operation by signal frequencies present in speech or music to which the receiver is exposed because of continuous association with the line. The guard channel uses frequencies outside the signal frequency band to oppose the operating effect of the signal frequency. In addition, this system inserts narrow-band elimination networks centered on the signal frequency in the voice path to limit the effect of signal power when this is present on the line. These networks are switched out at all other times.

5.3.2 *In-Band SF Signaling (Intertoll Applications)*

Fig. 15 is a photograph of the latest transistorized SF signaling unit (for intertoll trunks) in the Bell System. For comparison purposes, Fig. 16 shows the earlier electron tube model. Occupying one-third of

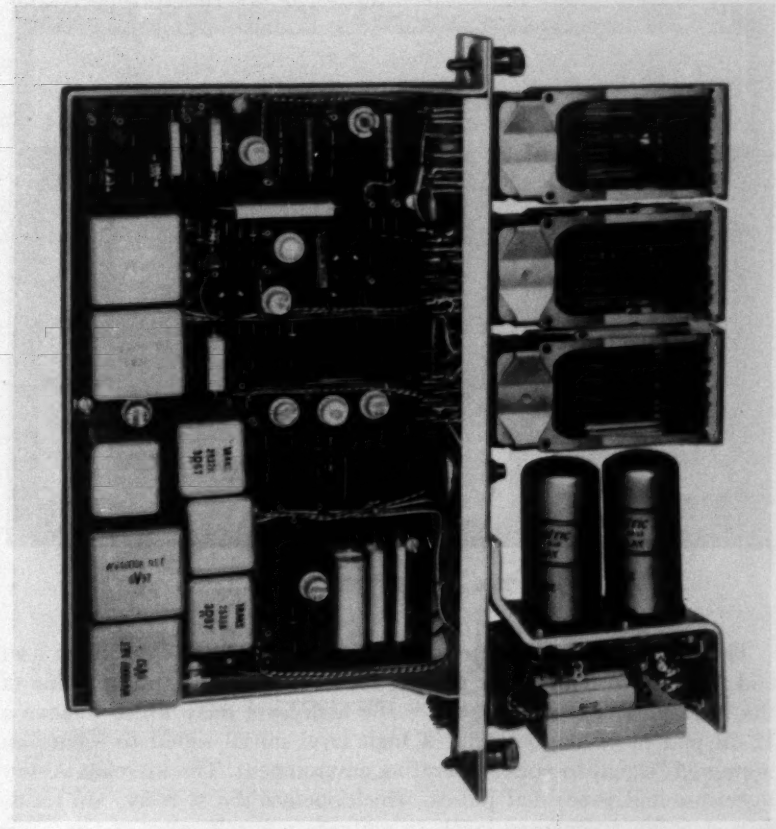


Fig. 15 — The E2B SF signaling unit.

the space and using one-third the power of its electron tube predecessors, the transistorized model employs miniaturized components, printed wired cards and appropriate solid state circuitry and techniques. Wire-spring relays are used where economics dictated their choice. The wedding of new and old concepts is proving to be a most happy match.

Fig. 17 indicates in block diagram form the more important features of the SF signaling unit. On the office side, six leads interconnect the signaling unit and associated trunk relay circuit, the E and M leads for signaling and the send and receive voice circuits of the signaling unit and the associated four-wire branches of the line facility, generally a carrier telephone system.

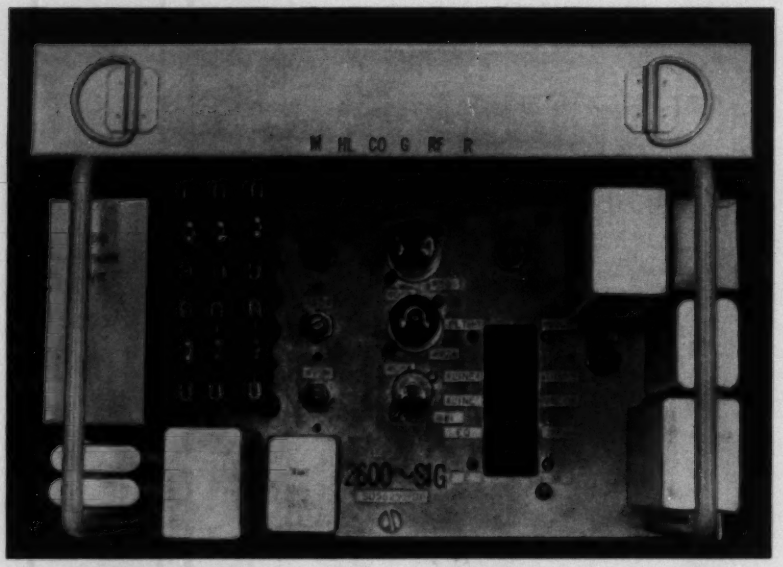


Fig. 16 — 2600-cycle SF unit (electron tubes).

The keyer relay M operates and releases from signals on the M lead and alternately removes or applies 2600 cycles to the transmit line of the facility. The M relay operates the high-level relay HL to remove a 12-db pad in order to permit a high level initial signal to secure an improved "signal-to-noise" operating environment. The HL relay is slow to release and hence dial pulses, which operate the M relay, are transmitted at an augmented level. In addition a cutoff relay, CO, operates to short-circuit any noise which may be present from the office side of the circuit. A bridged retardation coil with center-point grounded is also provided to drain off longitudinal currents that may arise in the office equipment and possibly interfere with signaling. The M lead will accept and the SF unit will transmit dial pulses at speeds from 8 to 12 pulses per second with per cent breaks ranging from 46 to 76 per cent.

The receiving portions of the SF unit include a voice amplifier, appropriate band elimination networks, and a signal detection circuit. The voice amplifier's primary function is to block any noise or speech present in the office equipment from interfering with the operation of the signal detector and also to make up for the insertion loss of the SF unit in the receive speech path.

The signal detector circuit includes an amplifier-limiter, a signal-to-guard network, appropriate half-wave rectifiers, a DC amplifier and a pulse-correcting circuit, the output of which operates a relay to repeat signals to the E lead of the trunk relay equipment.

The receiver sensitivity is −29 dbm for four-wire line facilities and −32 dbm for two-wire line facilities, the additional 3 db being for loss introduced by band elimination filters which are required when two different frequencies are used.

The signal-guard network provides the necessary frequency discrimination to separate signal and other than signal (guard) voltages. By combining the voltage outputs of the signal and guard detectors in opposing polarity, protection against false operation from speech and noise is secured. The efficiency of the guard feature is shifted between the dialing and talking conditions to secure optimum over-all operation.

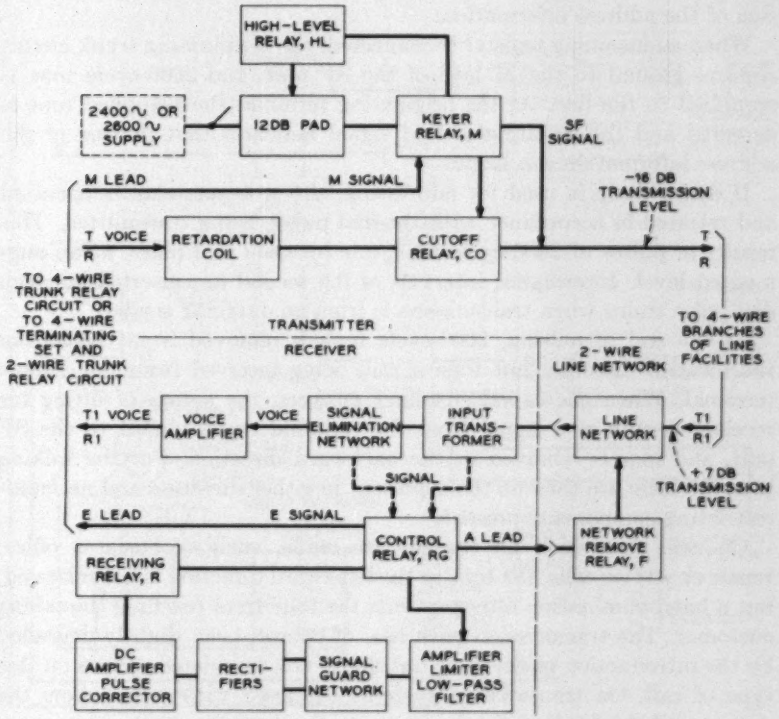


Fig. 17 — Basic elements of SF signaling circuit.

To understand the signaling unit operation, assume that a trunk equipped with 2600-cycle SF units is in the idle condition. A 2600-cycle tone (−20 dbm referred to the zero transmission level point, 0 db TL) is present in both directions of transmission.

A seizure signal originating at one terminal results in a change in state on the originating M lead from ground to battery. As a result, the keyer relay M operates and removes 2600-cycle tone from the sending line. The loss of tone is detected at the distant SF signaling unit and the resulting action changes the E lead from an open to a ground condition. The grounded E lead results in a request for connection of an incoming register or sender. Since there may be a slight delay in securing a register, the trunk circuit changes the M lead state at the terminating office from ground to battery, resulting in the removal of 2600-cycle tone from the line facility, from the terminating to the originating office. This results in a signal on the originating office E lead to delay the transmission of the address information.

When an incoming register is connected, the terminating trunk circuit restores ground to the M lead of the SF unit, and 2600-cycle tone is reapplied to the line. At the originating terminal the reapplied tone is detected and the resulting E lead signal indicates that pulsing of the address information can begin.

If dial pulsing is used for addressing, the M keyer relay is operated and released in accordance with the dial pulses being transmitted. This results in pulses of 2600-cycle tone, one for each dial pulse, at an augmented level. Interdigital intervals of 0.6 second are inserted between dial pulse trains when transmission is from an outgoing sender.

At the end of pulsing, 2600-cycle tone is removed from the line in the forward direction but tone is still being received from the distant terminal. When the called customer answers, the action of lifting his receiver results in a signal from the terminal trunk circuit to the SF unit, and tone is removed in the backward direction. For the talking interval there are then no tones present in either direction and no band-restricting components present.

On calls for which no charges are made, such as business office, repair or service calls, the tone in the backward direction is not removed, but a band-elimination filter prevents the tone from reaching the calling customer. The transmission path has, of course, been slightly degraded by the introduction of the filter but this is not considered serious on this type of call. On transmission systems equipped with compandors the presence of the backward-going tone may reduce the compandor cross-talk and noise advantage.

5.3.3 *Out-of-Band SF Signaling (Intertoll and Exchange Applications)*

Certain carrier systems include built-in signaling equipment (Types N, O, and ON) using a separate frequency slot outside of but adjacent to the voice band. In the Bell System a frequency of 3700 cycles is used for such signaling systems. Fig. 18 outlines the basic features.

During the trunk-idle condition the 3700-cycle frequency is present in both directions of transmission and trunk control and status signals are transmitted by interrupting the tone in similar fashion to that already described for in-band SF systems. Since the signaling path is outside of the voice band no provision is required for protection against voice operation. In addition, compandors are not affected by the tone, and signaling, if required, can take place during the talking condition.

The out-of-band system in use at the present time provides only for E and M connections to associated trunk circuits. Circuits have been

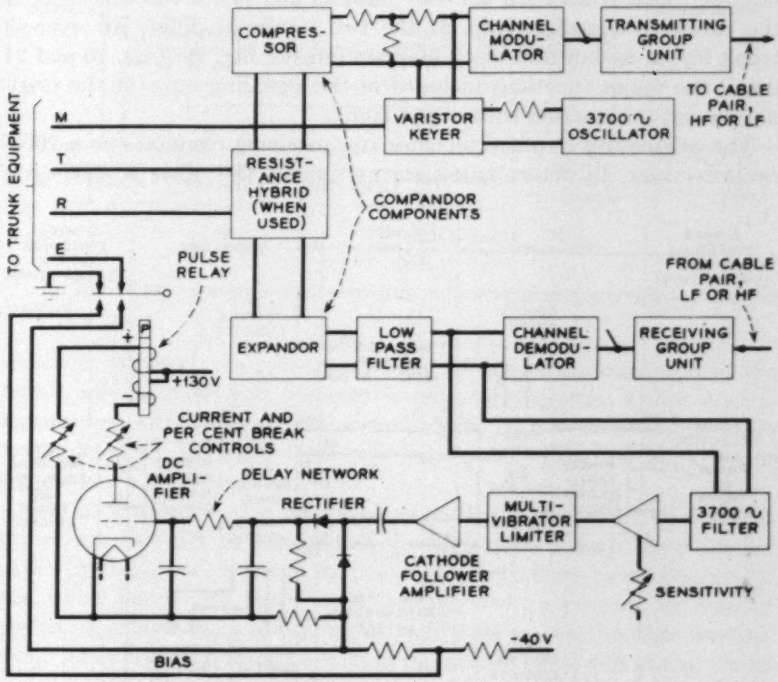


Fig. 18 — N-1 carrier signaling system (out-of-band).

designed for converting E and M lead signals to loop signals when required.

5.3.4 *In-Band SF Signaling; Revertive Pulsing (Exchange Applications)*

For application in the exchange plant a new series of transistorized signaling units makes it possible to adapt loop signaling trunks to short haul carrier systems. The SF units provide loop-signaling-reverse-battery supervision toward central office switching equipment and in-band AC signaling toward the line. These units also include the four-wire terminating sets required for converting between the four-wire line facility and the two-wire loop. The terminating sets are suitable from a transmission standpoint for exchange, tandem and toll-connecting trunks.

Since revertive pulsing, as indicated earlier, is a three-state signaling system, a brief description of its adaptation to AC signaling will be of interest. Such trunks are one-way circuits and hence the functions of the revertive signaling units at the two terminals differ. An over-all trunk layout is shown in block diagram form in Fig. 19. Figs. 20 and 21 detail the major functions included in the signaling units at the originating and terminating ends of the trunk.

The originating terminal includes two receiving circuits: one a 2600-cycle receiver, to detect trunk status signals, the other a 2000-cycle

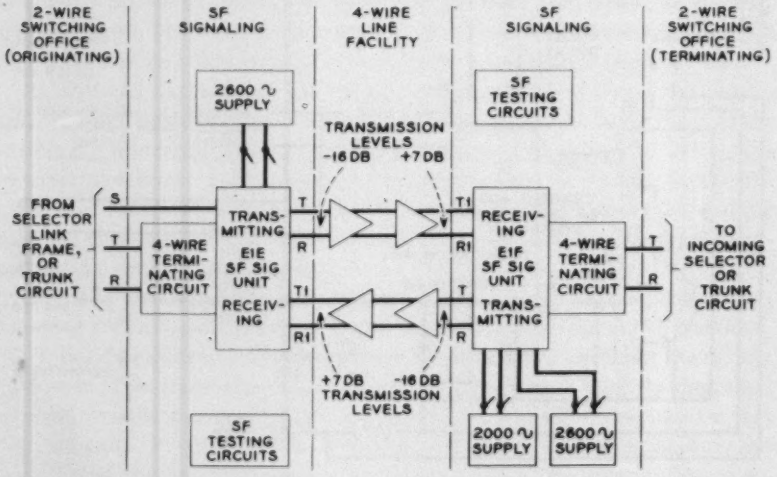


Fig. 19 — Circuit layout using SF signaling unit for revertive pulsing.

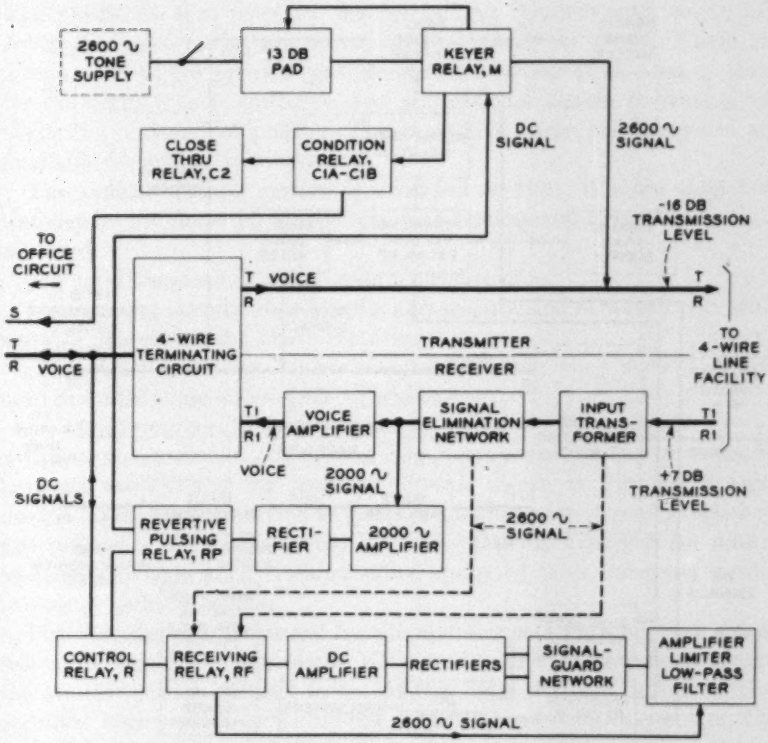


Fig. 20 — Basic elements of SF signaling unit — originating terminal revertive pulsing.

receiver, to detect the revertive pulses. At the terminating end of the trunk only a 2600-cycle receiver is required to detect trunk control signals but two transmitters are provided. The revertive pulses are transmitted by keying a 2000-cycle oscillator, while the trunk status signals key a 2600-cycle oscillator.

The bandwidth of the 2000-cycle channel for the revertive pulses is quite wide in order to accomodate the high-speed signals. As indicated earlier, pulsing speeds up to 32 pulses per second are used. Because of the wider bandwidth, care must be exercised to prevent extraneous circuit transients from appearing as legitimate pulses. Another requirement for revertive pulsing circuits is the need for fast round-trip transmission times. Since the revertive pulses are generated at the terminating end of the trunk and are detected and counted at the originating

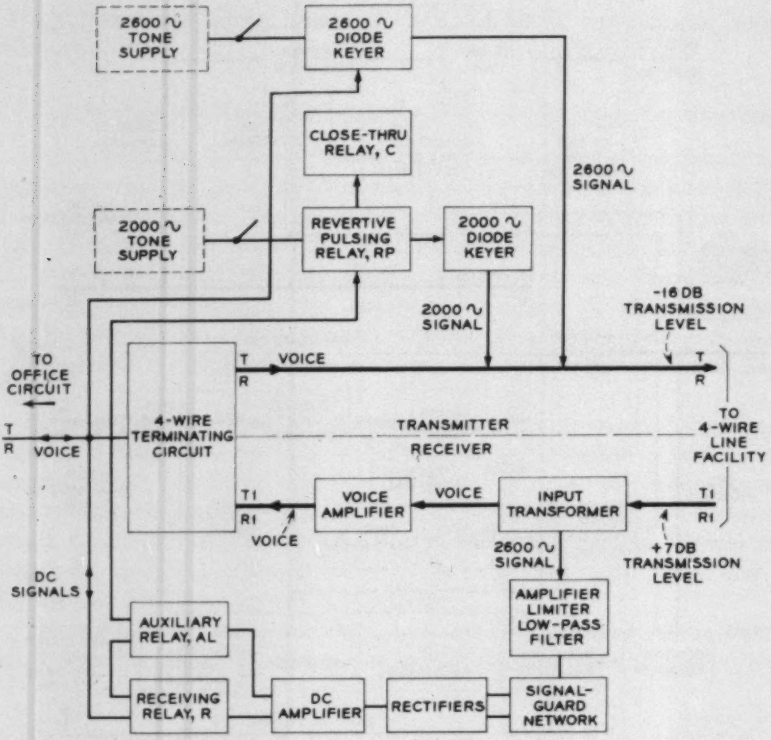


Fig. 21 — Basic elements of SF signaling unit — terminating terminal revertive pulsing.

terminal, a signal to stop the panel office elevator at the terminating end must be transmitted as soon as the counting circuit at the originating end is satisfied.

5.3.5 *Multifrequency Pulsing (MF)*

Multifrequency pulsing accomplishes the address function with AC signals, in both the toll and exchange plant. This system transmits digits by combinations of two, and only two, of five frequencies. Additional signals are provided by combinations using a sixth frequency. Table IV shows the multifrequency codes. The six frequencies are spaced 200 cycles apart, from 700 to 1700 cycles inclusive. Each combination of two frequencies represents a pulse, and each pulse represents a digit.

The pulses are sent over the regular talking channels and, since they are in the voice range, are transmitted as readily as speech. There are fifteen pairs of frequencies possible from the group of six; ten of them for the digits 0 to 9, inclusive, and one each for signals indicating the beginning and end of pulsing. The remaining three possible pairs are available for special requirements.

The multifrequency system is arranged so that, if more than two frequencies are detected by the receiving equipment, a reorder signal is returned.

The multifrequency pulsing system consists of:

(a) signaling current supply units and distribution arrangements with suitable protection and alarm features;

(b) signal transmitters, either manual multifrequency keysets or dial system multifrequency outpulsing senders;

(c) signal receiver connected to the incoming sender or register.

One arrangement of the system components for pulsing on a trunk between manual and dial system offices is shown in Fig. 22. A local manual or dial system or toll switchboard equipped with multifrequency key pulsing is shown connected by a direct trunk arranged for multifrequency pulsing to a crossbar office equipped with incoming multifrequency pulsing senders.

The first signal transmitted by the operator is a gate opening signal call a KP signal. Receipt of the KP signal at the distant end prepares the multifrequency receiver to accept the digits that are to follow. The operator begins sending the address digits by pressing one button for each digit. Following the last digit, she presses the ST button to indicate the end of pulsing. In addition to informing the distant sender that no more pulsing signals are coming, the operation of the ST key disconnects

TABLE IV — FREQUENCIES AND DIGIT CODES FOR MF PULSING

Digit	Code	Frequencies
1	0 + 1	700 + 900
2	0 + 2	700 + 1100
3	1 + 2	900 + 1100
4	0 + 4	700 + 1300
5	1 + 4	900 + 1300
6	2 + 4	1100 + 1300
7	0 + 7	700 + 1500
8	1 + 7	900 + 1500
9	2 + 7	1100 + 1500
0	4 + 7	1300 + 1500
KP	2 + 10	1100 + 1700
ST	7 + 10	1500 + 1700

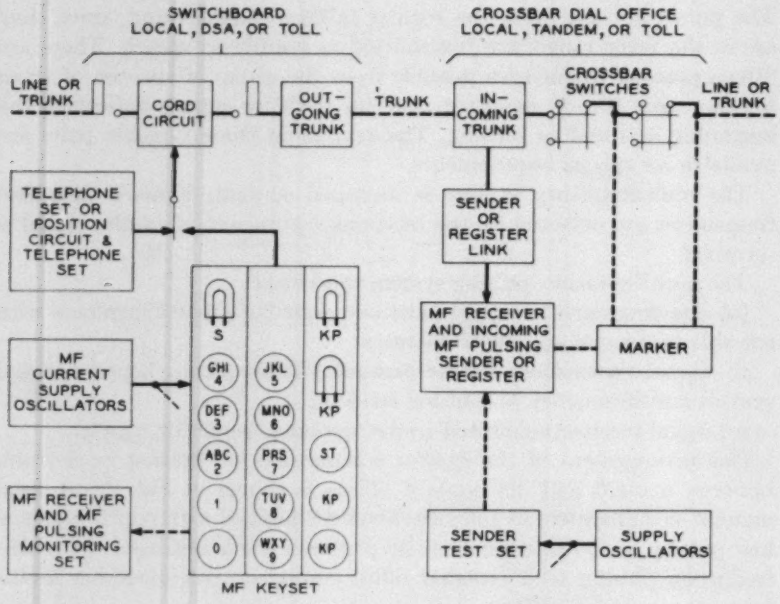


Fig. 22 — Plan of MF pulsing system.

the keyset from the cord, reconnects the telephone set under control of the talk key and extinguishes the KP and sender lamps.

If the operator inadvertently presses two buttons simultaneously, the MF receiver at the distant terminal will detect a variation from the two-out-of-six pattern, and a reorder signal is transmitted by the distant end. Given this information, the operator releases the connection and starts the call again. Fig. 23 shows a typical MF receiver plan.

When the receiver is in the pulsing condition the two signal frequencies applied to the input are amplified or limited depending upon the received signal power. The frequencies are selected by two appropriate channel filters and detected. The rectified voltages in turn fire hot-cathode gas tubes and in turn operate two channel relays. The channel relays when operated will ground two out of five leads to the incoming sender register circuit. When the received signal is ended, the signal present circuit of the receiver will indicate the end of the pulse to the register circuit, permitting the register to advance to the next register bank awaiting receipt of the next digit. Fig. 24 shows a multifrequency receiver equipment unit.

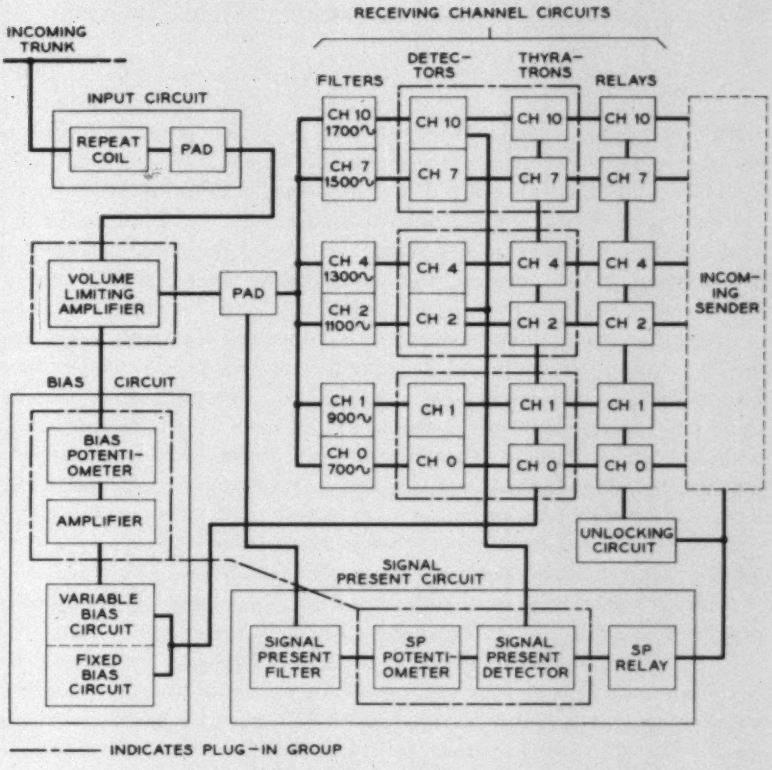


Fig. 23 — MF receiver plan.

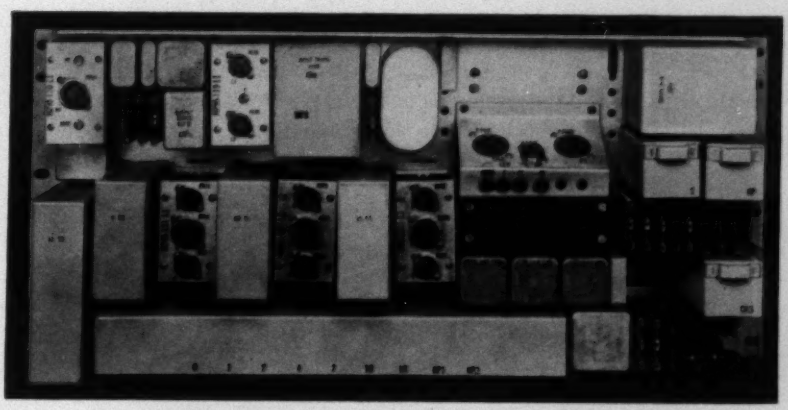


Fig. 24 — Multifrequency receiver unit.

VI. CONSIDERATIONS INVOLVED IN INTERNATIONAL SIGNALING

The extension of dial service to the United Kingdom and Europe introduces some interesting new signaling problems. Arising principally from the fact that there are some important differences between the switching philosophy and signaling techniques of the British and European telephone systems as compared to those of the telephone systems in this country, the signaling problem is further complicated by unique requirements imposed by new transmission systems. As has been noted in previous sections, such factors, while influencing the form of signaling systems during the evolution of telephone facilities in this country, have been subject to reasonably close coordination. However, up to recently there has been no need for coordination between North American systems and those used in Europe. Since the latter have independently reached an advanced state of the art, it is not surprising that rather complicated signaling problems now arise in attempting to convey switching information between hemispheres. These problems are reviewed below, together with the solutions that are being evolved.

International traffic over radio systems and submarine cables is at present handled by switchboard operators stationed at the circuit terminals. Transatlantic traffic presently employs "ringdown" operation; that is, an operator wishing to connect a calling customer to an international point signals an operator at the distant terminal by applying a ringing signal. The originating operator passes the wanted number verbally and the terminating operator completes the connection to the called customer. At the end of the call the originating operator again rings the receiving terminal and both operators pull down their connections. The signal used for ringdown operation from North America to Europe is a 500-cycle tone modulated by a 20-cycle tone; for calls from Europe to North America a 1000-cycle tone modulated by 20 cycles is used. Thus it can be seen that the international signaling function with ringdown operation is relatively simple.

Increased international business and improved transmission facilities have stimulated an increase in overseas telephone traffic. In order to utilize existing trunks more efficiently operator dialing is being applied initially to international circuits. There is the possibility of customer dialing at a later date. In this first step toward improved operations the originating operator dials the called customer's number to set up the call directly without the assistance of an operator in the terminating country. This type of operation is now in service between the U.S. mainland and Hawaii and Alaska, with operators dialing over radio

circuits as well as submarine cables. More recently, it has been applied to circuits between the mainland and Puerto Rico.

Dialing to European countries raises problems of compatibility between switching systems, signaling systems and operating techniques. The basic objectives of conveying information for trunk control, status information and addressing are the same for circuits connecting the United States with other countries as they are for circuits within the United States, but differences in signaling and switching philosophies create areas where compromises are necessary in the intersystem signaling arrangements. These differences are discussed below.

6.1 *New Information Requirements*

United Kingdom and European practices require that additional information be encoded, transmitted and interpreted by the associated signaling systems, as follows:

6.1.1 *Terminal and Transit-Seize Signals*

These signals are coded connect or seizure signals to indicate whether the incoming call should terminate in or switch through the country. In the Bell System such information is decoded by the switching system from address data.

6.1.2 *Language Signals*

These are digit signals (one through nine) indicating the service language which must be used when assistance operators are called in on the connection. This digit precedes the address information and is stored in the incoming relay or trunk circuit. When a ring-forward signal is received from the originating terminal, the stored language digit together with the operator selection signal will determine the operator to be called in on the connection. Such signals are not required in the Bell System.

6.1.3 *Acknowledgment Signals*

1. *Terminal and transit "proceed-to-send" signals* are used to indicate that the proper register is attached and ready to receive digits. The equivalent in the Bell System is the "start-dial" signal, except that only one signal is required.
2. *Digit reception acknowledgment signals* are used to inform the originating register that each individual digit has been received and that

transmission of the next digit can start. Digit acknowledgment signals are not used in the Bell System.

3. *Number-received signal* is a signal to indicate reception of the complete address. This signal is not required in the Bell System.

4. *Release-guard signal* is transmitted to the originating terminal to indicate that the clear-forward (disconnect) signal has been fully effective at the incoming end. It serves to protect the circuit from reseizure before the circuit has been completely disconnected. In the Bell System this protection is provided by timing arrangements at each terminal and no special signal is required.

6.1.4 *Operator Selection Signals*

Assistance operators handling European international telephone calls are designated as "Code 11" or "Code 12" operators. These operators are called in by transmission of distinctive signals and they perform the following services. A "Code 11" operator is an assistance operator who performs the usual functions of an incoming operator in manual service. A "Code 12" operator is a delayed-ticketing or suspended-call operator. When a particular "Code 12" operator is desired a call number is added and follows the "Code 12" signal.

6.1.5 *Typical Call*

To illustrate the use of these signals assume a call originated at London for Rome via Paris and Frankfurt. The initial seizure signal will be transmitted from the outgoing register in London to the incoming circuit in Paris as a "transit seize" signal. This signal will cause a "transit" register to be attached at Paris. When the register is ready to receive information a "transit-proceed-to-send" signal is returned to London. London will then transmit the country code digits representing the country of destination. The reception of this information at Paris will cause the register to return an acknowledgment of the digits to London and select a circuit to Frankfurt. A "transit-seize" signal is now transmitted by the Paris register to the Frankfurt terminal, and the speech path is closed through and the Paris register is dismissed.

The Frankfurt register, when connected, will return a "transit-proceed-to-send" signal to London, and London again transmits the country code digits. The Frankfurt register acknowledges reception of the country code, selects a route to Rome, transmits a "terminal-seize" signal to Rome, closes through the speech path and releases.

At Rome a terminal register is connected and a "terminal-proceed-to-

send" signal returned to London. The outgoing register in London now transmits a language digit and the national number of the called customer, followed by an "end-of-pulsing" signal. Each digit transmitted from London to Rome is acknowledged by transmission of a signal from Rome to London and, when the complete address has been received, the Rome register returns a "number-received" signal to the London operator. The outgoing register in London is dismissed and the speech path closed through after the "end-of-pulsing" signal is transmitted.

At the end of the call the connections are released at each switching center by disconnect and release guard signals transmitted in turn from Rome to Frankfurt, Frankfurt to Paris, and finally Paris to London.

The type of signaling described above is often referred to as "end-to-end" signaling. In contrast to this method the Bell System employs "point-to-point" signaling. In the latter all of the information concerning a call is transmitted from the originating point to the succeeding switching center, following which the sender at the originating point is dismissed to handle other calls. Signals are not transmitted between intertoll switching centers to acknowledge reception of the address information. Where calls require switching through several centers the address information includes an area code directing the call to the proper destination and when the final trunk is selected the area code information is dropped and only the customer's local address transmitted.

Point-to-point signaling results in reduced holding time for originating office senders, permits the use of different signaling and pulsing systems on each of the trunks comprising the complete connection (hence increased flexibility), and provides a transmission advantage since the signals are regenerated at each switching point. Hence, signaling arrangements for each trunk need only operate in the environment of that trunk and not that established for the entire connection. Finally, from an equipment viewpoint it is not necessary to provide two types of incoming registers for the handling of either transit or terminal traffic.

6.2 *One-Way Versus Two-Way Operation of Trunks*

As discussed in Section 4.5, Bell System long-haul intertoll trunks are generally operated on a two-way basis with exchange trunks being operated one way. In the United Kingdom and Europe one-way operation of trunks is the usual case. Two-way systems require symmetrical signaling terminals. The additional costs for symmetrical two-way operation as compared with unsymmetrical one-way operation are small for continuous-type signaling plans, but can be relatively expensive for spurt-

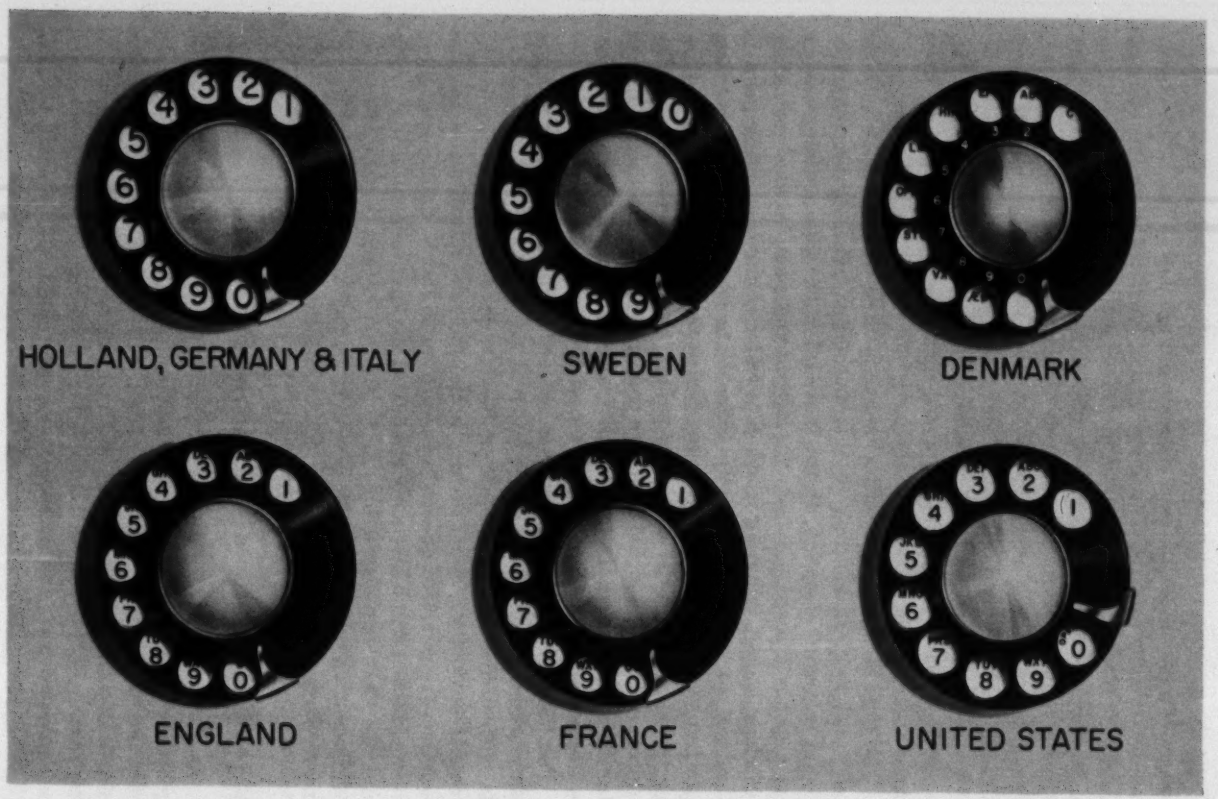


Fig. 25 — Comparison of dials.

type signaling arrangements, since additional logic and memory functions are required.

Two-way operation is essential for international trunks between North America and the United Kingdom and Europe because of the high costs of the transmission facilities and the desire to secure maximum trunk efficiency. Time zone differences increase the importance of being able to initiate calls in either direction at will.

6.3 *Differences in Dials*

Variations in the lettering and numbering of dials creates additional problems in international signaling. Fig. 25 indicates several of the arrangements used in the United Kingdom and Europe and also the United States dial. Dial differences are not serious with operator dialing, since bulletin information can be issued to operators handling calls to international points reflecting the necessary conversions. However, with customer dialing the problem is serious, since directory listings would need to be quite involved to cover the variations. A possible solution would be to adopt dials with all-numeral numbering, with the same numbers in respective holes for all countries.

Fig. 25 also points out differences in dials now using numerals only. For example the Swedish dial has the numeral 0 in the position where the numeral 1 is usually found on other dials. On the Swedish dials the digits run from 0 to 9 and on dials of other countries the digits run from 1 to 0. In this latter case the dial difference can be cared for by mechanical translation if it proves to be necessary.

6.4 *Numbering Plans*

Although the numbering plans of the various countries differ, it is not expected that this will influence the design of signaling systems. However, they will affect the associated switching systems, since the number of digits that customers must dial will affect the digit capacity of registers or senders handling the call.

6.5 *Charging for International Calls*

Allocation of charges on international calls is not primarily a signaling problem, but may at sometime require additional information signals.

6.6 *Signaling Techniques*

In recent years the desire of European countries and the United Kingdom to interconnect with each other has highlighted the necessity for

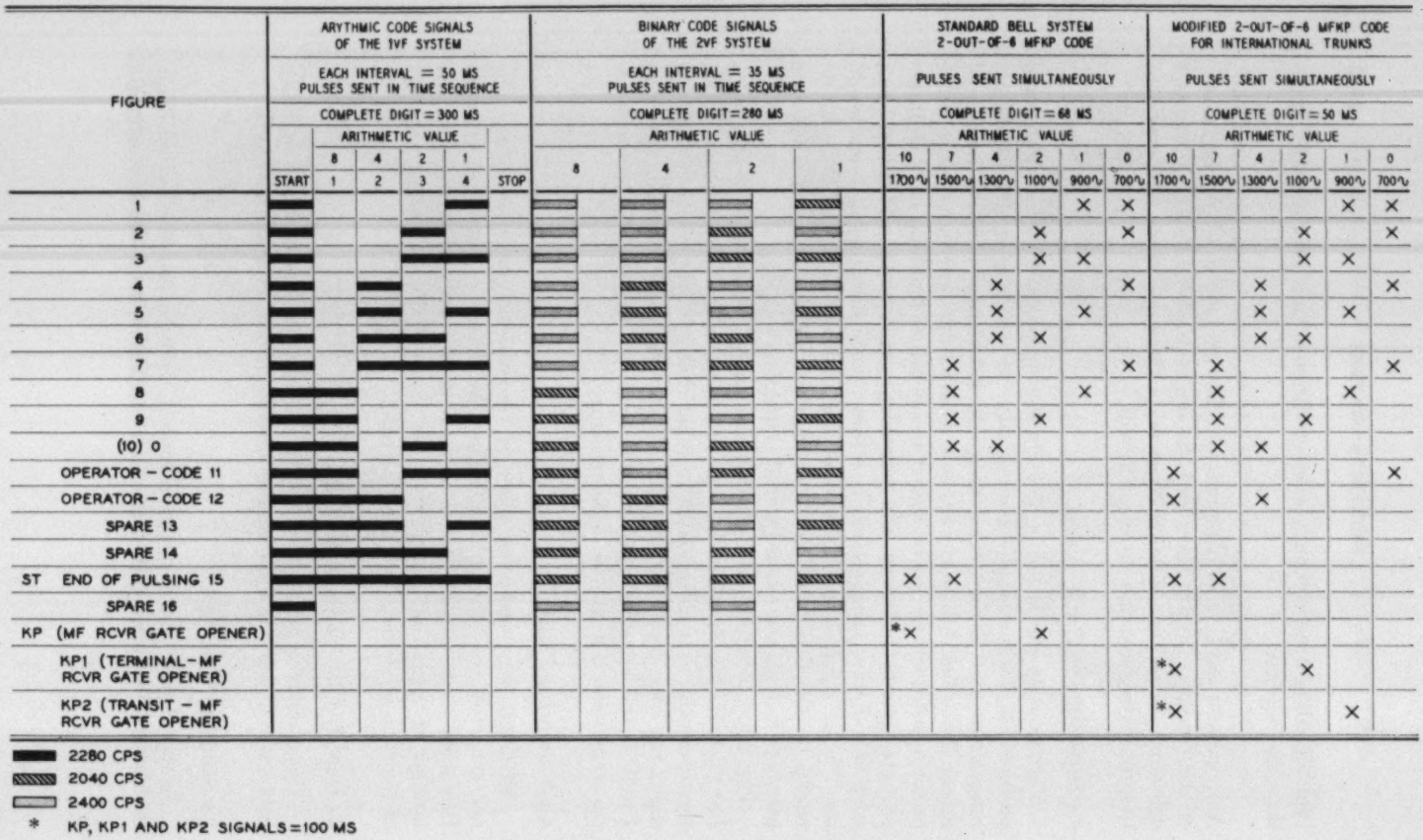


Fig. 26 — Comparison of codes for address information.

coordinating signaling systems in some fashion or agreeing to use one common system. The CCITT,* an international body of telephone administration representatives, has recommended two signaling proposals referred to as the single-frequency (1VF) and two-frequency (2VF) plans. Both systems use "spurt" signals as contrasted with "continuous" signals used in the Bell System.

In both of the "spurt" signaling systems there are signals present on the line facilities only during the intervals when it is required to transmit trunk control, status or address information. In the idle or busy conditions no tones are present whereas in continuous signaling systems, such as the Bell System SF signaling system, a low-level tone is present continuously during the idle condition. Usually no tone is present in the busy condition.

With spurt systems, signals such as connect, disconnect, on-hook, off-hook, etc. are transmitted by spurt signals comprising different combinations of frequencies and time intervals. Acknowledgment signals used to assure the two terminals of the successful transmission of the different bits of supervisory and pulsing information are also transmitted as spurts.

In addition to information and signal encoding differences in trunk control and status signals, European systems differ with respect to methods of transmitting address information. The CCITT standards were therefore proposed based on arythmic and binary-coded signals. The arythmic code, associated with the 1VF system, is quite similar to that used in telegraph systems. It comprises a start element, four signal elements and a stop element. The four signal elements provide 16 different combinations. (Telegraph codes usually employ five signal elements between the start and stop elements.) The binary code, associated with the 2VF system, comprises four elements separated each from the next by a silent interval, each element consisting of the transmission of one or the other of the two signaling frequencies. Here again, 16 different signal combinations are available. Fig. 26 shows both the arythmic and binary codes.

It is obvious that the above intersystem trunk control, status and addressing methods are markedly different compared to those used in the Bell System. In view of this, the interconnecting of switching systems on opposite sides of the Atlantic is a problem. Possible solutions offer the choice of either adopting the standard of one side or the other with suitable conversion arrangements on one side to accommodate the adopted standard to its foster parent or, alternatively, formulating a

* Comité Consultatif International Télégraphique et Téléphonique.

compromise arrangement having desirable features of both Bell and European standards, with conversion arrangements at each end. As noted below, the latter alternative has been selected, with the decision having been influenced by the transmission requirements to be discussed.

6.7 *Transmission*

To provide for economically expanding the number of available trunks between the United States and the United Kingdom and Europe, two plans are being implemented. First, the bandwidths of the available channels are being slightly reduced by spacing the channels at 3-kc intervals in place of the original 4-kc spacing, thus increasing the total number of channels to 48. Second, a new transmission facility, the Time Assignment Speech Interpolation (TASI) system is being employed. As many as 36 of the 3-kc channels may be associated with the TASI system.

A TASI system assigns a speech channel to a talker only when a channel is required, and when a channel is not required it is prepared to switch that channel to another talker requiring a channel. This makes it possible to use 36 channels to serve 72 talkers, since it has been found that a speech channel is in use less than 50 per cent of the time that a call is in progress. Thus, with reduced channel frequency spacing and use of TASI, a total of 84 trunks is made available from the original 36 channels. Actually, the number of trunks available will be somewhat less than the maximum due to the need for channels to provide services that cannot be handled on TASI-derived circuits.

The basic philosophy of TASI precludes the use of any signaling system which uses continuous tones, since such tones would prevent proper interpolation of channels. It is therefore necessary to devise a signaling system compatible not only with the systems used in the terminal countries served by the cables, but also with TASI-derived trunk operation. This problem has been the subject of intensive study, and discussions have been held jointly between Bell System and the overseas telephone administrations concerned. A plan of operation has been tentatively agreed upon as described below.

6.8 *Solution*

Basically the plan proposes that all international trunks be operated on a two-way basis and that the frequencies comprising the trunk control and status signals be 2400 and 2600 cycles, either separately or combined, to provide the necessary signal distinctions. For the trans-

mission of the address information and to reach special information operators it has been agreed to use a slightly modified form of the Bell System's multifrequency key pulsing (MFKP) system in both directions.

Figs. 26 and 27 compare the trunk control, status and address signals proposed for TASI-derived trunks between North American and European terminals with those normally used in the toll plant of the Bell System and those recommended by the CCITT for use on international trunks in the United Kingdom and Europe. It should be noted that certain CCITT signals are not used, but reliability has not been sacrificed since the principles of continuous-type signaling have been followed wherever possible (such as connect and disconnect signals) without seriously affecting the efficiency of TASI. The time intervals chosen for the various signals reflect the requirement that the availability of a TASI channel must be practically assured in every case. For the connect, start-dial, forward-disconnect and disconnect-acknowledgment signals, the nature of the signal and terminal recognition arrangements provide for signal lengths as required to assure the TASI channel. The signal on the TASI channel itself is the minimum time possible, consistent with reliable operation. The other signals are timed pulses, with part of each signal assigned to the securing of the TASI channel, if necessary, and the remainder of the signal for the use of the terminal equipment.

The North American terminals are to be equipped with signaling converters connected between the trunk circuit and the TASI channelizing equipment. The Bell System terminals will include two standard SF signaling units together with certain logic and memory relay circuitry. This equipment will convert the usual "continuous" types of signals between trunk and signaling circuits to signals compatible with the concepts of TASI operation and the signaling equipment at the European terminals.

In addition to the changes in signaling philosophy, switching equipment for international calls will include the following important features which also affect the signaling system:

1. Seizure of the TASI channel will be delayed until the outgoing register contains all the necessary address information. This is desirable in order to reduce the activity on the TASI channel.

2. The incoming register groups will be liberally engineered to assure availability of a register in a minimum time. This is again desirable to reduce the duration of the seizure or connect signal and hence reduce TASI activity.

3. When the TASI channel is available and the start-dial signal has been received from the distant terminal the address information will be

SIGNALS	STANDARD BELL SYSTEM SF SIGNALING FOR INTERTOLL TRUNKS (2-WAY OPERATION) FREQUENCY = 2600 CPS		CCITT 1VF SYSTEM FOR SIGNALING OVER INTERNATIONAL TRUNKS (1-WAY OPERATION) FREQUENCY = 2280 CPS		CCITT 2VF SYSTEM FOR SIGNALING OVER INTERNATIONAL TRUNKS (1-WAY OPERATION) FREQS. = 2400 CPS, 2040 CPS		SIGNALING FOR TRANSOCEANIC TASI DERIVED TRUNKS (2-WAY OPERATION) FREQUENCIES 2600 CPS, 2400 CPS	
	DIRECTION OF TRANSMISSION		DIRECTION OF TRANSMISSION		DIRECTION OF TRANSMISSION		DIRECTION OF TRANSMISSION	
	ORIG – TERM	TERM – ORIG	ORIG – TERM	TERM – ORIG	ORIG – TERM	TERM – ORIG	ORIG – TERM	TERM – ORIG
IDLE CONDITION	▷	▷	NO TONE	NO TONE	NO TONE	NO TONE	NO TONE	NO TONE
SEIZURE	NO TONE	▷	150 (TERMINAL); 600 (TRANSIT)	NO TONE	150, 250 (TERMINAL); 250, 150 (TRANSIT)	NO TONE	▷	NO TONE
DELAY DIAL	NO TONE	NO TONE	—	—	—	—	—	—
PROCEED TO SEND (START DIAL)	NO TONE	▷	NO TONE	150 (TERMINAL AND TRANSIT)	NO TONE	150 (TERMINAL); 150 (TRANSIT)	▷	▷
ADDRESS SIGNALS	MFKP	▷	ARYTHMIC CODE	NO TONE	BINARY CODE	NO TONE	MFKP	NO TONE
DIGIT RECEIVED	—	—	—	—	—	35 (TERMINAL); 35 (TRANSIT)	—	—
NUMBER RECEIVED	—	—	NO TONE	150	—	150, 150	—	—
CUSTOMER ANSWER	NO TONE	NO TONE	NO TONE	150, 150, 100	NO TONE	250, 150	NO TONE	850
CALLED CUSTOMER HANG-UP (ON HOOK, CLEAR BACK)	NO TONE	▷	NO TONE	600	NO TONE	150, 250	NO TONE	850
DISCONNECT-FORWARD (CLEAR FORWARD)	▷	▷	600, 100, 600	NO TONE	150, 350	NO TONE	▷ ▷	NO TONE
DISCONNECT – ACKNOWLEDGE (RELEASE GUARD)	—	—	NO TONE	600, 100, 600	NO TONE	350, 150	▷ ▷	▷ ▷
RING FORWARD (FORWARD TRANSFER)	70–130	NO TONE	150, 150, 100	NO TONE	350, 150	NO TONE	850	NO TONE
BUSY-REORDER-NO CIRCUIT (BUSY FLASH)	NO TONE	AUDIBLE TONE (30, 60, 120 IPM)	NO TONE	600	NO TONE	150, 250	NO TONE	AUDIBLE TONE (30, 60, 120 IPM) *

▷ CONTINUOUS TONE

DURATION OF PULSES INDICATED IN MILLISECONDS

* FOR CALLS TO UNITED KINGDOM & EUROPE, A BUSY FLASH SIGNAL MAY BE USED, EITHER ALONE OR ACCOMPANIED BY AUDIBLE TONE 850

Fig. 27 — Comparison of trunk control and status symbols.

transmitted without interruption and with signal intervals, both on and off time, of such duration as to guarantee the continuity of the channel.

4. Transmit and terminal-seize distinctions will be made by encoding the KP signal preceding the address signals.

5. A language signal consisting of an MF digit will, on terminal calls, be transmitted between the KP signal and the address to make it possible, when necessary, to call in an assistance operator. On transit calls the language digit will follow the two digits designating the international country code.

6. The terminating trunk circuit will be arranged to permit the calling-in of an assistance operator on a dial established connection when a ring-forward signal is received.

VII. FUTURE TRENDS

From the foregoing survey of the events and effects leading to the present state of the signaling art, it will be apparent that the signaling background has been deeply interrelated with the evolution of the telephone plant as a whole. Looking toward the future, one might logically conclude that signaling systems will continue to be closely allied with progress in the switching and transmission fields. Advances in the electronic component field and in related arts will undoubtedly contribute substantially to the technology of signaling, but future system trends will be tempered by the economic strains which inevitably occur in fitting tomorrow's plans to today's plant.

The foreseeable future, as indicated by trends already established, will bring more and more fully customer-controlled traffic — within the exchange plant as well as within the DDD network — and less operator-directed traffic. While this probably will not cause changes in signaling design requirements *per se*, other than to diminish the need for signals required only by operators, it will place more and more dependence on the reliability of signaling systems. Thus, one might expect a continued acceleration in the use of self-checking multifrequency signaling where a choice of engineering alternatives exists between this and other forms of signaling. By the same token, one might expect a lessening in the use of interoffice dial pulsing, either through gradual obsolescence of step-by-step central offices or through some form of senderization. There will probably be a decline in the use of PCI signaling as local manual switchboards disappear and as local office DDD sender modifications supersede, with MF signaling, the use of PCI signals to tandem offices. Further, as panel switching offices are replaced by more modern switching systems, there will be a corresponding reduction in the use of re-

vertive pulsing and a lessening of the tendency to employ this form of signaling between crossbar offices. Continued growth in the use of exchange carrier and small-gauge repeatered cable facilities will also promote an increasing use of AC signaling in the exchange plant. Thus, in the foreseeable future one might expect the emergence of multifrequency signaling as the basic standard for signaling between offices, with other methods of signaling falling by the wayside.

Beyond the more obvious trends lies the possibility that the future may bring to fruition some of the signaling systems and services which are now only in the experimental or introductory state. Of interest among these are pushbutton signaling systems wherein the customer's conventional telephone dial is displaced by ten (or perhaps more) pushbuttons, each of which is capable of generating a unique signal, similar to the multifrequency signaling system. Such systems require that register (or sender) operation of some form be present in the central office switching systems to receive and act upon the signals generated by the customer's pushbuttons. Since, as mentioned earlier, a sender-type central office serves as a buffer between the customer and interoffice signaling systems, customer pushbutton systems probably will not influence directly present-day interoffice signaling systems. However, it is possible that an indirect effect may result, since by injecting pushbutton register-senders into non-common-control offices, the potential use of such senders to transmit other than dial pulses to connected switching systems may become attractive.

The use of AC types of pushbutton operation will undoubtedly stimulate intercustomer signaling, wherein pushbutton signals originated at one station will be used to control devices or switching operations at a distant station. Careful integration of these over-all signaling systems with the interoffice signaling systems will be required, however, to ensure satisfactory transmission and to avoid mutual interference.

Wider use of other systems as yet in the experimental stage may in the future result in substantial changes in interoffice signaling. For example, high-speed data links carrying signaling information for large groups of trunks might become attractive; there also exist new signaling possibilities such as digital switching and transmission systems employing pulse code modulation techniques. In the latter system one or two bits are added to each of the digital "words" used to transmit a quantized speech sample, and these bits convey the signaling information. Such signaling systems may in the future be competitive with present-day systems; they have the capability of transmitting signaling information at high speed. In this connection, electronic switching systems,

which can efficiently handle high-speed interoffice signaling, may bring about wide use of high-speed frequency division systems (characterized by multifrequency systems) or time division, as characterized by PCM or similar systems.

Looking toward the more distant future, one discerns the possibility that switching systems will be able to capitalize on higher and higher signaling speeds between switching offices, that transmission systems will be available to handle these speeds economically and that message services (such as, for example, data communications over the telephone plant) having high calling rates and short holding times will place a high premium on rapid switching and signaling. It is a matter of conjecture whether increased signaling speed will be obtained by frequency-division systems, by serial transmission or by one or the other in combination with multiplexing techniques. One also anticipates the possibility that the direct dialing network will be broadened to include new areas, new countries and perhaps new dimensions — and that this and other new services will probably require new techniques and the transmission of additional signals conveying new information.

Finally, taking the long look toward sheer fantasy (the inevitable goal of most predictors of the future) one is struck by the thought that it is necessary to reach way out to insure that a prediction of future communication possibilities will indeed be fantastic. Stopping just short of this, one wonders if Alexander Graham Bell's historic telephone summons to his assistant, Mr. Watson, might one day be repeated into some exotic communication network, and that "Watson" will be summoned to the telephone more quickly than his predecessor was able to reach Mr. Bell's side. If switching engineers can devise switching facilities to do this, and if transmission engineers will do their part, signaling engineers will do their best to see that Mr. Watson is summoned, wherever he is.

BIBLIOGRAPHY

1. Scudder, F. J. and Reynolds, J. H., Crossbar Dial Telephone Switching System, B.S.T.J., **18**, 1939, p. 76.
2. Sheppard, H. A., A Signaling System for Intertoll Dialing, Bell Lab. Rec., **18**, 1940, p. 337.
3. Abraham, L. G., Busch, A. J. and Shipley, F. F., Crossbar Toll Switching System, A.I.E.E. Trans., **63**, 1944, p. 302.
4. Dahlbom, C. A., Horton, A. W., Jr. and Moody, D. L., Applications of Multifrequency Pulsing in Switching, A.I.E.E. Trans., **68**, 1949, p. 392.
5. Korn, F. A. and Ferguson, J. G., The No. 5 Crossbar Dial Telephone Switching System, A.I.E.E. Trans., **69**, Pt. 1, 1950, p. 244.
6. Caruthers, R. S., Kahl, W. E. and Pedersen, L., The N-1 Carrier Telephone System, B.S.T.J., **30**, 1951, pp. 1; 418.
7. Clark, A. B., The Development of Telephony in the United States, A.I.E.E. Trans., **71**, Pt. 1, 1952, p. 348.

8. Myers, O., Common Control Telephone Switching Systems, B.S.T.J., **31**, 1952, p. 1086.
9. Yokelson, B. J., New Multifrequency Receiver, Bell Lab. Rec., **32**, 1954, p. 221.
10. Comité Consultatif International Télégraphique et Téléphonique, XVIIth Plenary Assembly, Geneva, October 4–12, 1954, Vol. V, Signaling and Switching.
11. Weaver, A. and Newell, N. A., In-Band Single-Frequency Signaling, B.S.T.J., **33**, 1954, p. 1309.
12. Horton, A. W., Jr. and Vaughan, H. E., Transmission of Digital Information over Telephone Circuits, B.S.T.J., **34**, 1955, p. 511.
13. Adam, A. O., Crossbar Tandem as a Long Distance Switching System, B.S.T.J. **35**, 1956, p. 91.
14. Joel, A. E., Jr., An Experimental Switching System Using New Electronic Techniques, B.S.T.J., **37**, 1958, p. 1091.
15. Pennoyer, D. H., Automatic Number Identification and Its Application to No. 1 Crossbar, Panel and Step-by-Step Offices, B.S.T.J., **37**, 1958, p. 1295.
16. Bullington, K. and Fraser, J. M., Engineering Aspects of TASI, B.S.T.J., **38**, 1959, p. 353.
17. Green, E. I., Evolving Technology of Communication, Elect. Engg., **78**, 1959, p. 470.
18. Vaughan, H. E., Research Model for Time-Separation Integrated Communication, B.S.T.J., **38**, July 1959, p. 909.
19. Dahlbom, C. A., Transistorized Signaling System: Engineering Aspects, Bell Lab. Rec., **37**, 1959, p. 254.
20. Schenker, L., Pushbutton Calling with a Two-Group Voice-Frequency Code, B.S.T.J., **39**, 1960, p. 235.
21. Newell, N. A., DX Signaling, Bell Lab. Rec., **38**, 1960, p. 216.

The E6 Negative Impedance Repeater

By A. L. BONNER, J. L. GARRISON and W. J. KOPP

(Manuscript received April 14, 1960)

The E6 repeater is a low-cost, transistorized, voice-frequency, two-way repeater for use in the exchange area plant, including use in trunks between local offices and between local and toll offices. It permits both types of trunks to be operated at low net loss, while meeting the impedance requirements placed on trunks between local and toll offices. These features are possible because of a unique method of repeater design, in which the impedance-matching function and the gain function are allocated to separate portions of the repeater and are separately adjustable.

The impedance-matching function is accomplished by means of a line building-out network, which permits any cable pair associated with the repeater to be built out and made to match the impedance of the toll office. The gain unit presents the same image impedance, which is essentially independent of gain adjustment. Changes in gain are controlled by two resistive, adjustable networks — one in the series-connected negative impedance converter, and the other in the shunt-connected converter. In previous designs of negative impedance repeaters, complex impedance networks were required for adjustment of both gain and impedance match.

The E6 repeater also incorporates novel mechanical features, with transistorization resulting in low power drain and less heat dissipation in the telephone central office. Power is derived from the existing 48-volt central office battery.

TABLE OF CONTENTS

I. Introduction 1446
1.1 The E23 Negative Impedance Electron Tube Repeater 1446
1.2 The E6 Negative Impedance Transistorized Repeater 1447
II. Systems Applications 1447
2.1 General Description 1447
2.2 Principles of Operation 1448
III. Design Objectives 1451
IV. Theory 1452
4.1 General 1452
4.2 The E23 Repeater 1452
4.3 The E6 Repeater 1453
4.4 Line Building-Out Networks 1468

V. Performance Characteristics of the E6 Repeater 1483
5.1 Impedance and Return Loss 1483
5.2 Load Capacity and DC Stability 1485
5.3 Susceptibility to Battery Noise 1485
5.4 Lightning Surge Protection 1486
VI. Apparatus Description of the E6 Repeater 1486
VII. Transmission Testing Equipment 1491
7.1 Introduction 1491
7.2 54A Transmission Measuring Set 1492
7.3 54B Test Stand 1493
7.4 54C Return Loss Measuring Set 1493
VIII. Engineering Considerations 1495
IX. Systems Results and Comparison with E23 Rpeeater 1497
X. Acknowledgments 1504
References 1504

I. INTRODUCTION

1.1 *The E23 Negative Impedance Electron Tube Repeater*

One of the methods used to reduce the transmission losses in exchange area cable circuits employs E23* negative impedance repeaters. These repeaters,[1] first installed in 1954, represented a considerable advance over the series-type negative impedance repeaters[2] then in use. The improvement was accomplished by the addition of a shunt-type converter to the existing series-connected negative impedance element. Impedance discontinuities introduced into otherwise uniform loaded circuits and the resulting undesirable reflection effects were thereby avoided. The effectiveness of this technique is attested by the fact that over one million E repeater units have been installed in the Bell System since 1954.

Installation of the E23 repeater is relatively complex, since the negative impedance of the repeater must be closely matched to the cable impedance in order to give the desired gain and return loss characteristic. The matching is accomplished by strapping components of complex networks included in the repeater. The required strapping varies with (a) the amount of gain required, (b) the cable facility with which the repeater is associated and (c) the length of cable end-section to which the repeater is connected. Network connections for a particular condition are specified by extensive strapping charts. This is a time-consuming operation on the initial installation and on subsequent changes, either for gain or end-section adjustment.

A basic limitation on the application of the E23 repeater to telephone offices is that it is an electron tube device requiring both a 130-volt DC

* An E23 repeater is made up of an E2 series-type repeater and an E3 shunt-type repeater connected in a bridged-T arrangement.

plate supply and a 48-volt DC filament source. The 48-volt supply is almost universally available in telephone offices, but the necessity for a 130-volt plate supply has limited the use of this repeater to offices where such voltage supplies exist, or where the cost of installing a reliable 130-volt source could be justified.

1.2 *The E6 Negative Impedance Transistorized Repeater*

In reviewing the disadvantages of the E23 repeater it was apparent that by use of transistors in the gain portion of the repeater the need for a 130-volt supply could be eliminated, operating costs be drastically reduced and the annoying problem of heat dissipation in large installations be largely removed. Design studies also indicated that definite advantages would result if it were possible to associate with the repeater a wideband matching network that would match the complex impedance of the associated cable facility to the image impedance of the converter section of the repeater. By thus simplifying the impedance matching requirements on the converter, it appeared possible to develop a negative impedance repeater requiring only resistive networks for gain and return loss adjustment. Thus the concept of the E6 repeater was evolved.

II. SYSTEMS APPLICATIONS

2.1 *General Description*

The application of negative impedance repeaters to the exchange area plant is dependent on the type of service performed by the trunk facilities with which they are to be associated. If a repeater introduces an impedance discontinuity into an otherwise uniform loaded circuit, a substantial amount of energy will be returned to the sending end as "echo." The quality of transmission is related to the magnitude of this echo and to the delay introduced by the transmission facility before the echo reaches the talker's ear. As a result, when the delay is small, as in the case of an interoffice trunk for short-haul service, a relatively large amount of echo can be permitted. On toll-connecting trunks, however, the delay may be large and only a small amount of echo can be allowed. Since the gain obtainable from a series-type negative impedance repeater is dependent on the amount of negative impedance inserted in series with the trunk, it follows that the resulting echo is proportional to the gain of the repeater. Application of the series-type negative repeater is, therefore, limited, except at low gains, to interoffice trunk use. The E23 repeater, because of its better return loss, can be used with cor-

respondingly higher gains. The E6 repeater is even more flexible in specific applications inasmuch as its maximum gain is largely determined by the structural return losses of the connecting circuits rather than by repeater characteristics. Later in the discussion, the performance of the E23 and E6 repeaters will be compared in detail.

2.2 *Principles of Operation — E23 and E6 Repeaters*

The block diagram of Fig. 1 illustrates the principles of the E23 repeater. The single block labelled E23 CONVERTER contains the negative impedance controlling features of the repeater. In the diagram, the E23 repeater is shown at the local office end of a toll-connecting trunk circuit. Inserted between the line and the toll office is a standard impedance compensator used to improve the match between the office and the loaded cable facility. Branching out from the repeater, but included as part of the series and shunt converters comprising it, are two variable, complex impedance networks. These networks are designed so that the impedance of the converter over the voice-frequency band, for a particular gain adjustment, closely approximates the impedance of the connected cable facility. Element values are obtained by strapping of passive elements selected from the 51 elements available within the repeater.

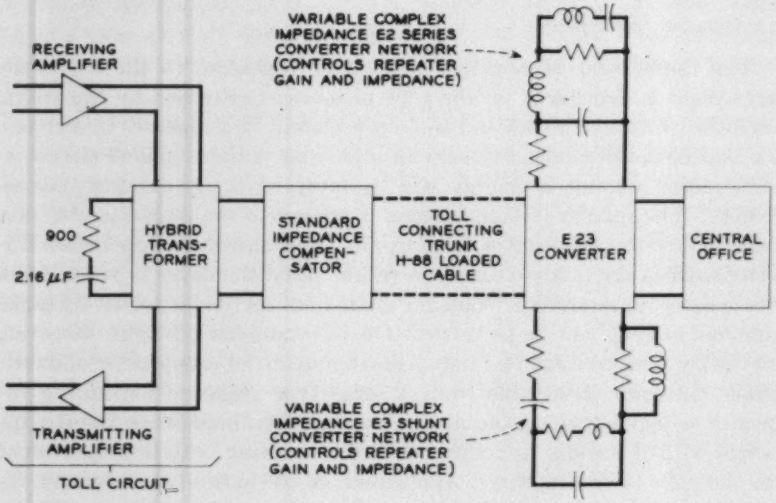

VARIABLE COMPLEX IMPEDANCE E2 SERIES CONVERTER NETWORK (CONTROLS REPEATER GAIN AND IMPEDANCE)
RECEIVING AMPLIFIER
900
2.16 μF
HYBRID TRANS-FORMER
STANDARD IMPEDANCE COMPEN-SATOR
TOLL CONNECTING TRUNK H-88 LOADED CABLE
E 23 CONVERTER
CENTRAL OFFICE
TRANSMITTING AMPLIFIER
TOLL CIRCUIT
VARIABLE COMPLEX IMPEDANCE E3 SHUNT CONVERTER NETWORK (CONTROLS REPEATER GAIN AND IMPEDANCE)

Fig. 1 — E23 repeater installation; terminal use — toll-connecting trunks.

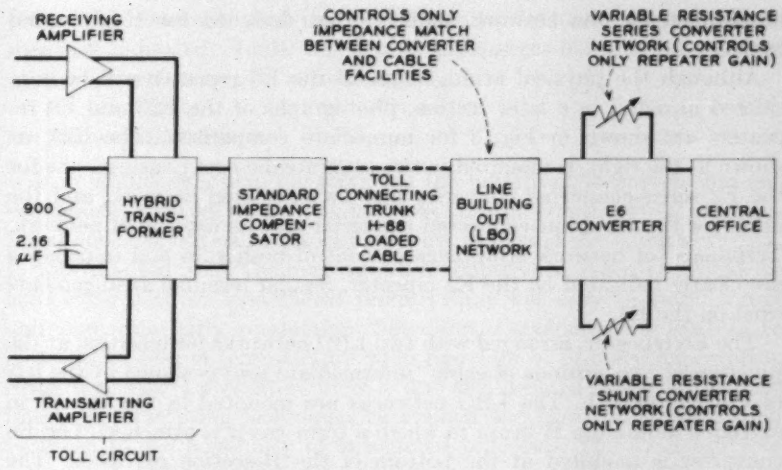

RECEIVING AMPLIFIER
CONTROLS ONLY IMPEDANCE MATCH BETWEEN CONVERTER AND CABLE FACILITIES
VARIABLE RESISTANCE SERIES CONVERTER NETWORK (CONTROLS ONLY REPEATER GAIN)
900
2.16 μF
HYBRID TRANS-FORMER
STANDARD IMPEDANCE COMPEN-SATOR
TOLL CONNECTING TRUNK H-88 LOADED CABLE
LINE BUILDING OUT (LBO) NETWORK
E6 CONVERTER
CENTRAL OFFICE
VARIABLE RESISTANCE SHUNT CONVERTER NETWORK (CONTROLS ONLY REPEATER GAIN)
TRANSMITTING AMPLIFIER
TOLL CIRCUIT

Fig. 2 — E6 repeater installation; terminal use — toll-connecting trunks.

For comparison purposes, the application of the E6 repeater to the same situation is shown in Fig. 2. It should be re-emphasized that the configuration of the two converters and the philosophy underlying the two designs are not the same. The complex impedance networks used in the E23 repeater have been reduced in the E6 repeater to simple adjustable resistances, and a wideband matching or line building-out network (LBO) has been added to match the impedance of the converter to that of the trunk. Inasmuch as the E6 repeater is situated at the local office end, an impedance compensator is still required at the toll office end to improve the match to the office impedance. The adjustment of the line building-out network of the E6 repeater is a simple operation involving the loosening or tightening of a number of screw connections. The gain of the repeater is similarly adjusted. This compares to strapping and soldering a number of terminals on the E23 repeater.

It should be pointed out, however, that the E6 repeater, although more efficient and easier to adjust, does not provide the wide flexibility of use inherent in the E23 repeater. A separate wideband matching network is required for each type of loaded facility to which it is to be connected, whereas the E23 repeater can approximate the impedance of a wide variety of circuits. Individual networks can, however, be provided containing only those elements required for a particular line facility. Fortunately, more than 85 per cent of existing exchange area circuits can

be handled by one network, namely, that designed for H-88 loaded cable.

Although the physical arrangement of the E6 repeater will be considered in detail in a later section, photographs of the E23 and E6 repeaters are shown in Fig. 3 for immediate comparison. The E23, as shown at the right, is mounted in two separate die-cast housings, one for the E2 series-connected converter and its associated network, and the other for the E3 shunt-connected converter and its associated network. Terminals for network strappings to control both gain and impedance are clearly indicated on the E2 repeater. Similar terminal arrangements exist on the E3.

The E6 repeater, arranged with two LBO networks for insertion at the junction of two sections of cable (intermediate use) is shown at the left of the photograph. The LBO networks are mounted in the top of an extruded aluminum H-beam to which a front cover is attached. The E6 converter is mounted at the bottom of the H-section extrusion. The

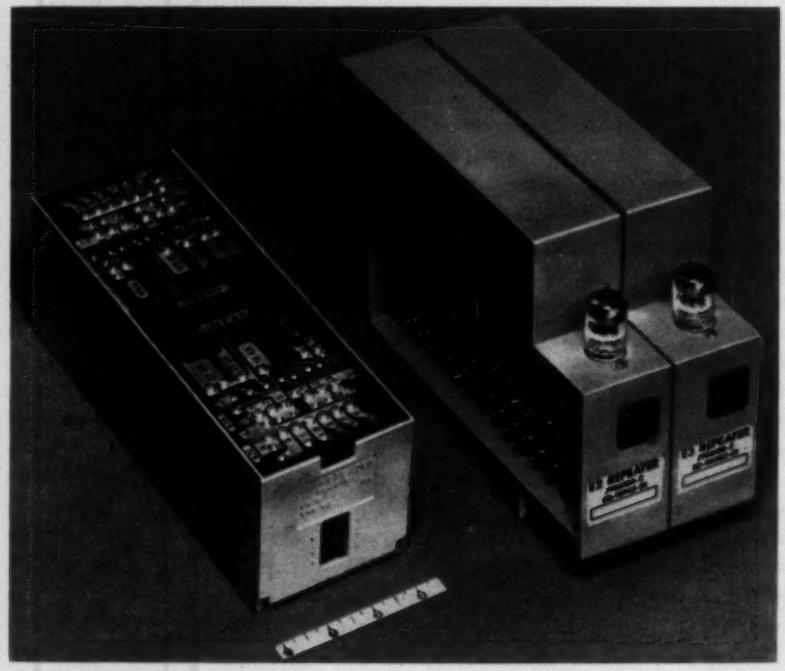


Fig. 3 — E6 repeater (left); E23 repeater (right).

screw-type terminals used on the LBO networks for impedance adjustment are indicated. Similar screw-type adjustment features are provided on the converter for gain adjustment.

III. DESIGN OBJECTIVES

In establishing a plan for achieving the important improvements and economies believed to be attainable by use of transistorized negative impedance converters in association with wideband matching networks, it was realized that the new negative impedance repeater would be competing with an established product that was easy to manufacture and free from early production difficulties. Considerations relating to cost and to ease of manufacture were therefore of primary importance. From a production point of view it was also deemed essential that the latest manufacturing techniques be applied and that full consideration be given to the possibility of automated processes. The availability of low-cost, long-lived transistors was essential to the realization of the desired economies. Since transistors are inherently low-power devices, protection had to be provided against the effects of (a) foreign potentials in the telephone plant due to lightning and induced surges from power circuits and (b) high potentials due to signaling. Such surges may occur from either tip and ring wires to ground or between tip to ring wires. Surge peaks greater than 600 volts are normally limited by carbon-electrode discharge gaps, so that any repeater must be designed to tolerate voltages up to this magnitude.

Electrical requirements were established to meet the projected needs of the Bell System exchange area plant. In recent years there has been the objective of reducing the losses in this area of the plant in order to improve subscriber-to-subscriber transmission. Today's objective for direct trunks is 4 db average, while 2 db to 3 db maximum is required for tandem or toll-connecting trunks. Broadly interpreted, two objectives were outlined for the repeater:

i. To provide a transistorized voice-frequency negative impedance repeater having series and shunt converters (in association with wideband matching networks) so arranged as to reduce circuit net losses in discrete adjustable steps by means of variable resistance networks and to present an image impedance equal to 900 ohms in series with 2.16 microfarads; i.e., 900 + 2.16 microfarads.

ii. To provide a wideband impedance matching network (line building-out network) capable of matching variable end section loaded cable facilities of different gages to the impedance presented by the converter section of the repeater under conditions of terminal or intermediate use.

IV. THEORY

4.1 *General*

Negative impedance repeaters such as the E23 and E6 are based on a bridged-T network configuration with mutual coupling. The standard bridged-T configuration for the case of positive impedance elements is shown in Fig. 4(a). The image impedance and insertion loss of the network are seen to be simply related to the impedances Z_A and Z_B of the series and shunt elements comprising it. Similar relations, indicated in Fig. 4(b), exist for the case of negative impedance elements, and the defining equations for image impedance and insertion gain are identical in form to the previous case except that substitution of negative impedance elements results in a gain instead of a loss. This is more apparent if the negative sign of the series arm is written as

$$Z_A \angle 180°$$

and the shunt arm as

$$Z_B \angle 180°.$$

4.2 *The E23 Repeater*

If a negative impedance repeater were to be inserted into a distortion-

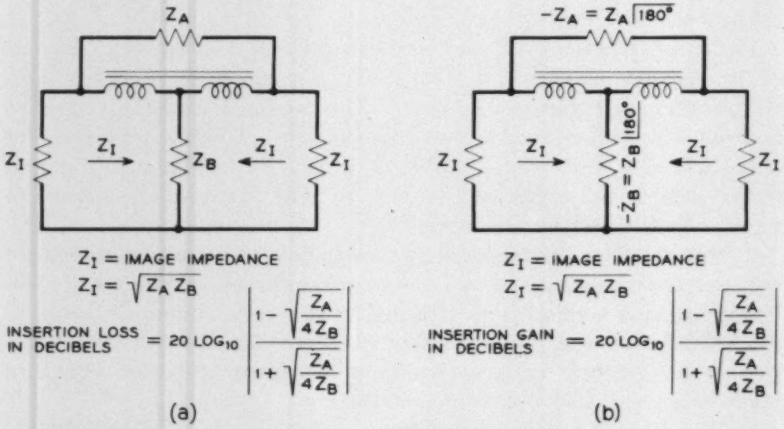


Fig. 4 — (a) Bridged-T positive impedance network; (b) bridged-T negative impedance network.

less transmission line, without introducing impedance discontinuities, the image impedance of the repeater would have to match the characteristic impedance of the line. In the E23 repeater this impedance match was established by properly relating the series and shunt elements of the equivalent bridged-T structure to the characteristic impedance of the line by a proportionality constant N. Assuming that the negative series impedance Z_A of Fig. 4(b) is equated to $NZ_0 \boxed{180°}$, and that the negative shunt impedance Z_B is set equal to $Z_0/N \boxed{180°}$, the image impedance of the repeater, Z_I, may be written

$$Z_I = \sqrt{Z_A Z_B} = Z_0,$$

where Z_0 = the characteristic impedance of the line.

Similarly, if the above expressions for Z_A and Z_B are substituted into the equation for insertion gain given in Fig. 4(b), the insertion gain may be expressed as

$$\text{insertion gain (in db)} = 20 \log_{10} \left| \frac{1 + N/2}{1 - N/2} \right|.$$

The above expression for insertion gain may be used to determine stability limits for such repeaters. If $N = 2$, the insertion gain is infinite and the repeater will sing. The inherent limitations on available gain from negative impedance repeaters was recognized at an early stage. In 1919, while presenting an historic account of the research and development work that led to the first successful hybrid-type telephone repeaters, Gherardi and Jewett[3] briefly described the condition for stability of the series-type negative repeater. Later, in 1931, Crisson[4] considered both series- and shunt-type negative impedance repeaters and presented a detailed account of the conditions that had to be imposed for realization of a stable circuit. The first practical negative impedance repeater was due to Merrill,[2] who developed a negative impedance electron tube repeater, the stability and operation of which could be accurately predicted.

4.3 *The E6 Repeater*

In the E6 repeater impedance-conversion, impedance-matching and gain-adjusting functions are, in contrast to the E23 repeater, allocated to separate portions of the repeater. All impedance-matching and impedance-adjusting features are accomplished in the line building-out networks and gain adjustment is relegated to the series- and shunt-type elements of the converter. As a result, the theory of operation of the

line building-out networks or of the converter may be considered separately once the impedance of the terminating cable as seen through the LBO network has been specified.

4.3.1 *Representation of Cable and LBO Network by a Lattice*

In the E6 repeater, certain portions of the LBO network are used — as the name "line building-out" implies — to build out fractional end-sections of cable to full-section value. The remaining portions of the LBO network consist of high- and low-frequency correcting networks used to match the impedance of the built-out cable to that of the converter.

The impedance and loss-frequency characteristic of such built-out and impedance-transformed cable can be represented by an equivalent lattice network as is shown in Fig. 5. The E6 repeater, arranged for terminal use, is shown in the upper portion of the figure. The image impedances appearing at each junction of the circuit are indicated. Considering only that portion of the circuit between the vertical broken lines, it can be seen that the LBO network, loaded cable facility and impedance compensator (nonrepeatered LBO network) have been replaced by an equivalent lattice having variable series and shunt impedance arms. Expressions defining the image impedance and insertion loss of an

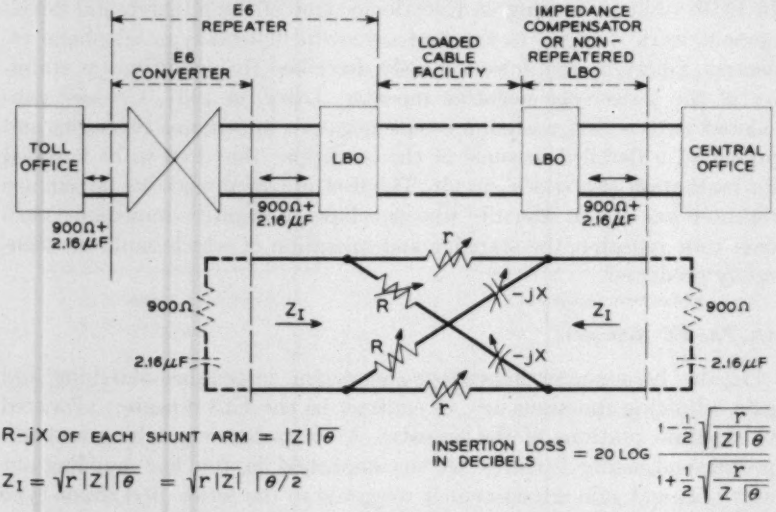


Fig. 5 — Equivalent lattice network of built-out section of loaded cable.

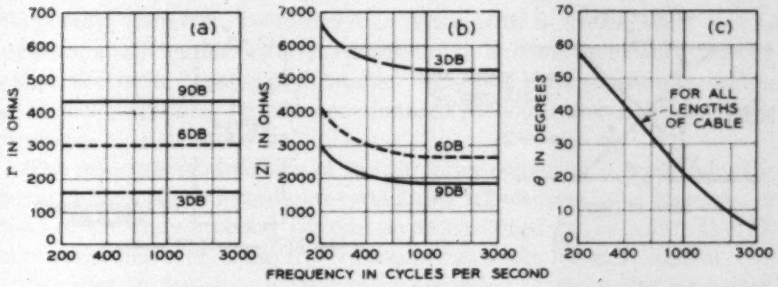


Fig. 6 — Approximate values of r and $Z \angle\theta$ for built-out 19-gage H-88 cable.

equivalent lattice are also shown. Although an impedance compensator, or nonrepeatered LBO network, is shown at the central office end of the trunk circuit, this network is usually omitted.

The magnitude of the series-arm impedance of the equivalent network of Fig. 5 increases with cable length, while the magnitude of the shunt-arm impedance decreases. Approximate values of the series- and shunt-arm impedances for lengths of cable having losses of 3, 6 and 9 db at 1000 cps are shown in Fig. 6. It is important to note that the angle θ specified for the shunt arm of the lattice is constant for all lengths of cable, varying only as a function of frequency.

4.3.2 *Lattice and Bridged-T Equivalents of E6 Converter*

Inasmuch as a lattice representation of the cable and LBO network has been defined, impedance requirements for both the series- and shunt-type converters of the repeater may now be established. In Fig. 5, assume that the E6 converter is replaced by an equivalent lattice. Assume further that the series and shunt arms of this lattice are equal in magnitude, but opposite in phase to those of the lattice equivalent of the built-out cable. Under these conditions, the image impedances presented by the converter will be maintained and the gain-frequency characteristic of the lattice will compensate for the loss-frequency characteristic of the built-out cable. The equivalent lattice network of the converter is shown in Fig. 7(a). Expressions for image impedance and insertion-gain are also shown.

Although the lattice representation of the E6 converter is a convenient tool for analysis, the bridged-T equivalent is more convenient for physical realization. In Fig. 7(b) the negative impedances required for the series and shunt arms of such a configuration are indicated. In Fig. 7(c)

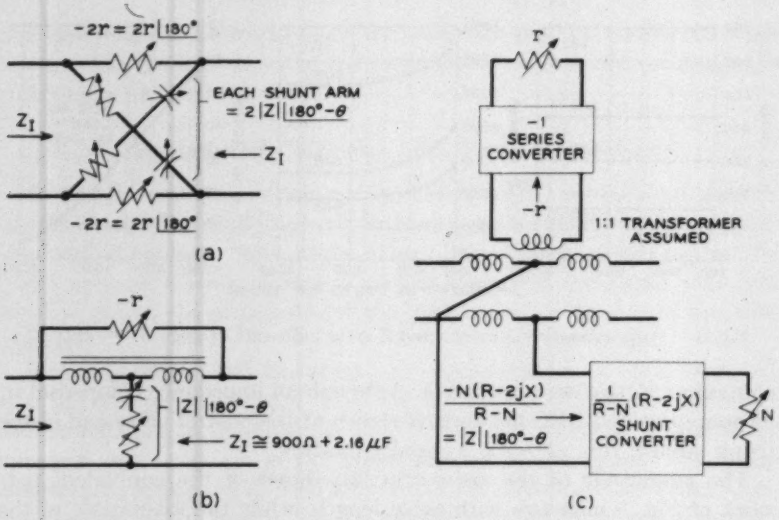


Fig. 7 — The E6 converter and equivalent circuits: (a) lattice network equivalent of E6 converter; (b) bridged-T equivalent of E6 converter; (c) E6 series and shunt converters.

the actual series and shunt converters are defined and shown in a balanced configuration employing a line transformer. It will be noted that the gain-adjusting networks associated with both converters are resistive. The series converter has a conversion factor of -1, while the shunt converter requires a conversion factor of $-N(R - 2jX)/(R - N)$. These conversion factors will be derived in discussion of the idealized E6 converter.

A more detailed schematic of the E6 converter is presented in Fig. 8. Only those circuit elements necessary for an understanding of converter operation are included. Furthermore, those elements which have no effect on idealized or theoretical operation are shown dotted. These include the simplified bias arrangements and transformers T2 and T3, which prevent longitudinal line voltages from disturbing the DC biases.

4.3.3 *Series Converter of the E6 Repeater*

The series-connected converter of the E6 repeater is, ideally, a four-terminal network having an input impedance that is the negative of the terminating resistances connected across its output. Such negative impedances, which are of the series or reversed-voltage type,[4] are open-circuit stable. Under certain restrictions that will be developed later,

they may, therefore, be inserted in series with a transmission line to produce amplification without rendering the line self-oscillatory. Specifically, the term "open-circuit stable" indicates that the line terminals of the converter may be left open-circuited without causing the converter to sing.

The principle of operation of the E6 series converter is essentially the same as that of the impedance-terminated E2 electron tube repeater, or that of the transistorized E2 prototype described by Linvill.[5] The E6 converter, however, uses only resistance networks for gain control.

In order to describe the theory of operation of the series converter, the two-transistor arrangement previously shown in Fig. 8 will be employed. In the actual converter, as will be seen later, a double-compound transistor arrangement involving four transistors is used, but the simplification in no way invalidates the theory. Further simplification is

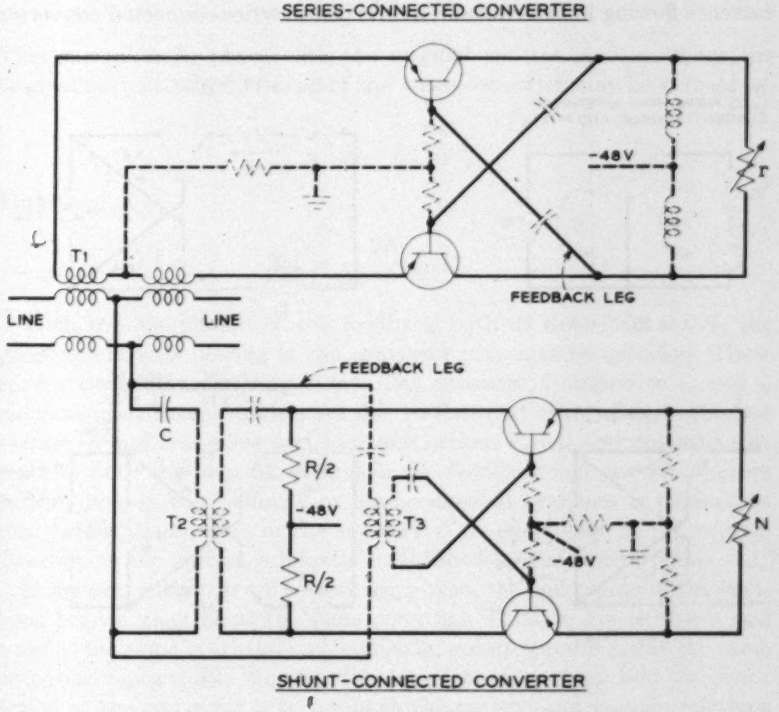


Fig. 8 — Simplified circuit of E6 converters with single transistors in each side of push-pull circuits.

achieved by neglecting biasing resistors, and by assuming that transistor alphas are equal to unity and that base and emitter resistances approach zero.

For the simplified schematics of Fig. 8, mesh equations could very easily be written and solved for the input impedance of the series or shunt connected converters. However, as soon as parasitic elements, cross-coupling or feedback capacitors and compensating or balancing elements are introduced, the mesh equations become unwieldy and difficult to interpret. If, however, a step-by-step description of the currents and voltages existing in the circuit is attempted, considerable simplification results.

In the usual mesh analysis, emitter and collector currents flowing in the circuit are assumed to have been established. Fig. 9(a) shows the direction of current flow in both emitter and collector circuits of a grounded-base transistor circuit. In Fig. 9(b), however, only the initial currents flowing in the emitter circuits of the series-connected converter

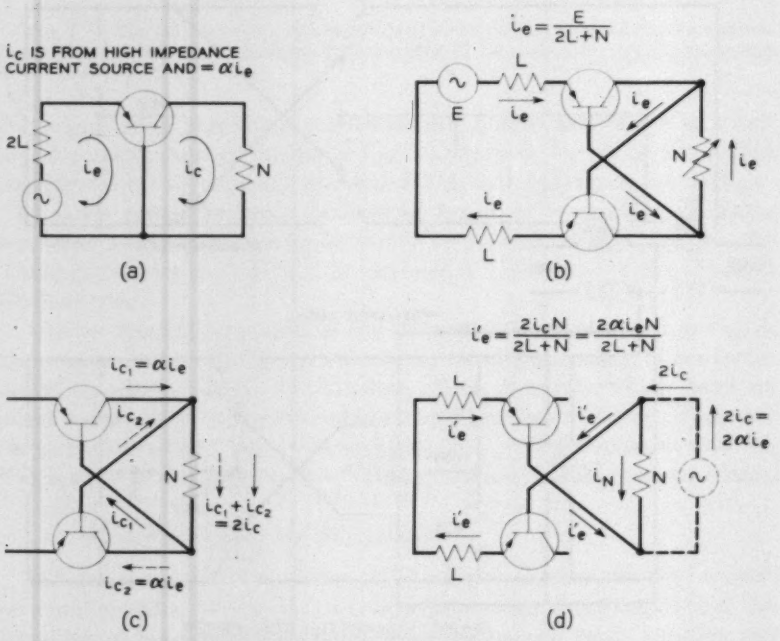


Fig. 9 — (a) Method of showing transistor currents; (b) emitter current due to line voltage E; (c) collector currents due to emitter current of (b); (d) positive feedback path completed through lines $2L$ to give i_e'.

due to application of the line voltage E are considered. As these currents flow, transistor action sets up collector currents in the collector portion of the circuit. In Fig. 9(c), only the collector currents flowing in circuit are shown. Each collector contributes a current equal to αi_e, and these combine in the adjustable resistive network N to give a total current of $2i_c$ or $2\alpha i_e$.

Considered from a slightly different point of view, the network of Fig. 9(c) may be replaced by that of Fig. 9(d), in which the current $2i_c$ is derived from an infinite impedance current source. In Fig. 9(d) the positive feedback path existing in the series-connected converter is completed through the line impedances, and the current $2i_c$ divides between these line impedances and the network N in inverse proportion to their magnitudes. Thus,

$$i_e' = \frac{2i_cN}{2L + N} = \frac{2\alpha i_eN}{2L + N}.$$

This current is in phase with the original emitter current. Since the original current was i_e, the $\mu\beta$ of the series converter may be defined as:

$$\mu\beta = \frac{i_e'}{i_e} = \frac{2\alpha N}{2L + N}.$$

For $\alpha = 1$,

$$\mu\beta = \frac{2N}{2L + N}.$$

With the completion of the feedback path as described above, the effective currents flowing in the converter may now be specified. These currents are shown in Fig. 10(a). The subscript T added to i_e and i_c indicates that these currents are due to the total effect of both the line voltage E and transistor and feedback action. Equal and opposite currents in both the network N and in the feedback legs cancel. The net current flow in the feedback or base-connected branches is reduced to zero, while the current in the network N is effectively in the reverse direction to the current originally established by the emitter alone.

If the net current in a feedback leg is zero, the end-points of the feedback branch must be at the same potential. Thus, in Fig. 10(b), A and B are at the same potential and, similarly, points C and D are at the same potential. Since this is the case, the net effect of feedback and transistor action in the converter is to establish the current and voltage relations shown in Fig. 10(b). If the voltage drop across network N is V, the voltage across the input terminals of the converter is also V, but is reversed

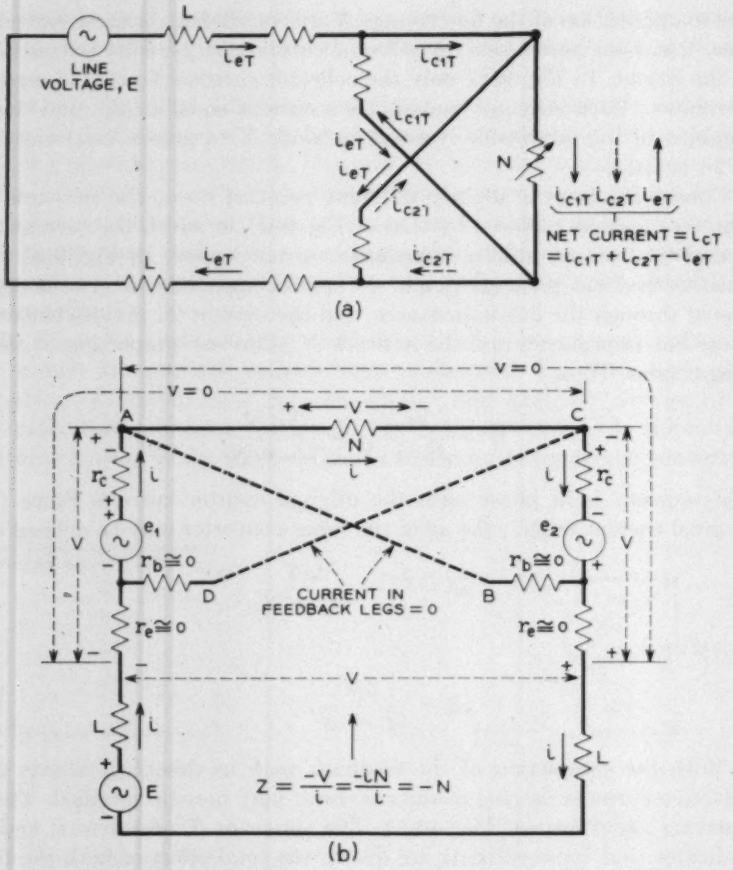


Fig. 10 — (a) Total currents in series converter due to line voltage and feedback; (b) voltages and currents in active series converter.

in sign. This voltage is in series with the line and, hence, the name "series or reversed voltage" type of converter.

In order for the voltages around the emitter-collector of the circuit to meet the requirements of Kirchoff's law, the four voltages indicated around the periphery must occur with the magnitudes and polarities as shown. The polarities of the voltages e_1 and e_2 in the collector circuits control the polarities of the voltage V across the transistors, since the effective collector voltage drop is small due to positive current feedback.

Considering the foregoing description of circuit operation from the standpoint of summing voltages around the circuit loops of Fig. 10(b), the input impedance of the series converter may be written by inspection, i.e.,

$$Z = -\frac{V}{i} = \frac{-iN}{i} = -N.$$

This impedance is closely approximated in the E6 series-connected converter for frequencies between 300 and 3000 cps and up to gains of approximately 13 db for the complete E6 repeater.

4.3.4 *Shunt Converter of the E6 Repeater*

Like the series converter, the shunt converter of the E6 repeater is a resistance-terminated four-terminal network. The shunt converter, however, does not present a resistive load to the line, but instead presents a complex impedance that is derived from a fixed phase-shift network included within the converter.

The shunt converter, as contrasted to its series counterpart, is of the shunt or reversed-current type.[4] It may, therefore, be placed in shunt across a transmission line without rendering the line self-oscillatory. Such a converter is short-circuit stable and, as a result, the input terminals of the device may be short-circuited without causing the converter to break into oscillation.

The basic principle of operation of the shunt converter is very similar to that of the series type. It should be noted, however, that the negative impedance of the shunt converter is obtained at the collectors of the converter rather than at the emitters. This is indicated in Fig. 11.

In developing the theory of operation of the shunt converter, the same step-by-step approach used in the series case will be found of value. In this instance, however, the derived voltages, indicated by the symbol V and shown around the outer periphery of the shunt converter circuit of Fig. 11, include the voltage drops across the phase-shift reactances $(-jX)$ as well as those across the collectors. It will also be noted that the voltage V across the input terminals of the converter is of the same polarity as that of the voltage developed across L2. This is the reverse situation from that of the series case. Furthermore, the total current i_3 flowing out of the converter terminals is seen to be in the reverse direction from that caused to flow by application of the voltage E to the line. This gives rise to the designation "reversed-current negative impedance".

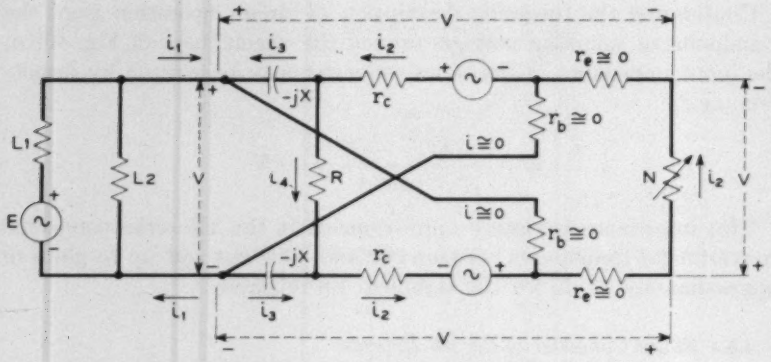


Fig. 11 — E6 phase-shift shunt converter.

As in the series case, the currents flowing in the feedback branches of the converter cancel and the relationship $V = i_2N$ can be written immediately. The other relationships needed for derivation of the negative impedance presented at the converter terminals follow from inspection of Fig. 11:

$$i_2 = i_3 + i_4,$$
$$V = i_3 2jX + i_4 R,$$
$$i_3 = -i_1.$$

Solving,

$$i_1 = \frac{V(R - N)}{-N(R - 2jX)}.$$

Since $Z = V/i_1$, the expression for the converter input impedance follows:

$$Z = \frac{-N(R - 2jX)}{R - N}.$$

This can also be written

$$Z = \frac{-N(R - 2jX)}{R - N} = |Z| \quad \underline{|180° - \theta}.$$

Were it not for the phase shifter, the expression for Z would be merely

$$Z = -N.$$

The more complex relationship, obtainable with the phase shifter, represents the desired impedance for the shunt converter of the E6 repeater. It is evident from inspection of this equation that the phase angle of Z is dependent solely on the magnitudes of R and $-2jX$ and is in no way dependent on N. The magnitude of N can thus be varied at will to adjust repeater gain without effect upon the phase angle of the resulting negative impedance. The impedance Z is closely approximated by the E6 shunt converter for gains as high as approximately 13 db for the complete E6 repeater.

The phase shifter of the shunt converter controls the impedance of the E6 repeater and permits impedance matching of the repeater to the impedance of the line as modified by the LBO network. The phase shifter also controls the gain-frequency characteristic of the repeater as a function of gain setting. These features of the phase-shift network become apparent when the series and shunt converters are combined in the standard bridged-T configuration.

If the input impedance of the series-connected converter is denoted as Z_A and that of the shunt converter as Z_B the following relations may, as has been previously indicated, be written:

$$Z_A = -N_1 \angle 0^\circ = N_1 \angle 180^\circ,$$

$$Z_B = \frac{-N_2}{R - N_2}(R - 2jX) = \frac{-N_2}{R - N_2}(R - jX)$$

$$= \frac{N_2}{R - N_2}\sqrt{R^2 + X^2} \quad \angle 180 - \theta_2,$$

where

$$\theta_2 = \tan^{-1}\frac{X}{R}.$$

Since the image impedance of bridged-T connected negative impedance converters is specified by the relation

$$Z_I = \sqrt{Z_A Z_B},$$

substitution of the derived expressions for Z_A and Z_B results in:

$$Z_I = \sqrt{\frac{N_1 N_2 \sqrt{R^2 + X^2}}{R - N_2}} \quad \angle \frac{-\theta_2}{2}.$$

The image impedance of the E6 converter must approximate 900 ohms

in series with 2.16 microfarads. This latter impedance may be expressed as:

$$Z_0 = R_0 - jX_0 = \sqrt{R_0^2 + X_0^2} \quad \underline{\frac{-\theta_0}{2}},$$

where

$$R_0 = 900,$$

$$X_0 = \frac{10^6}{2.16\omega} \text{ ohms}$$

and

$$\theta_0 = \tan^{-1}\left(\frac{X_0}{R_0}\right).$$

For Z_I to equal Z_0, the following equalities must hold:

$$R_0 = \sqrt{\frac{N_1 N_2 R^2 + X^2}{R - N_2}}$$

and

$$\tan^{-1}\left(\frac{X_0}{R_0}\right) = \tan^{-1}\left(\frac{X}{R}\right).$$

Except for the X^2 term appearing under the radical in the equality involving R_0, these relationships would be exact, and X^2 is small compared to R^2 at frequencies above 300 cycles. The degree of approximation that may be obtained is dependent on the proportion of N_1, N_2, R and X. These same factors determine the gain and gain-frequency characteristics of the converter and, as a result, both impedance and gain requirements must be considered in the selection of design values.

In order to derive an expression for insertion gain of the E6 converter, reference is again made to the general bridged-T network for which

$$\text{insertion gain (in db)} = 20 \log_{10} \left| \frac{1 - \sqrt{\frac{Z_A}{4Z_B}}}{1 + \sqrt{\frac{Z_A}{4Z_B}}} \right|.$$

Considering only the terms under the radical sign, and substituting expressions for Z_A and Z_B,

$$\sqrt{\frac{Z_A}{4Z_B}} = \frac{1}{2}\sqrt{\frac{N_1(R - N_2)}{N_2\sqrt{R^2 + X^2}}} \underline{\theta_2/2} = M \underline{\theta_2/2}$$

Insertion gain may, therefore, be written

$$\text{insertion gain (in db)} = 20 \log_{10} \left| \frac{1 - M \underline{/\theta_2/2}}{1 + M \underline{/\theta_2/2}} \right|$$

Taking the real and imaginary parts of $M \underline{/\theta_2/2}$, this expression becomes:

$$\text{insertion gain (in db)} = 20 \log_{10} \left| \frac{\left[1 - M \cos\left(\frac{\theta_2}{2}\right)\right] - jM \sin\left(\frac{\theta_2}{2}\right)}{\left[1 + M \cos\left(\frac{\theta_2}{2}\right)\right] + jM \sin\left(\frac{\theta_2}{2}\right)} \right|.$$

When the absolute value of this equation is taken, the expression for insertion gain reduces simply to:

$$\text{insertion gain (in db)} = 20 \log_{10} \sqrt{\frac{1 - \delta}{1 + \delta}},$$

where

$$\delta = M \left[M + 2 \cos\left(\frac{\theta_2}{2}\right) \right],$$

$$M = \frac{1}{2} \sqrt{\frac{N_1(R - N_2)}{N_2 \sqrt{R^2 + X^2}}}$$

and

$$\theta_2 = \tan^{-1}\left(\frac{X}{R}\right).$$

As indicated in this equation, the over-all gain of the bridged-T connected converter is dependent on the gain settings of the series- and shunt-connected units (N_1 and N_2) and on the phase-shift parameters R, X and θ_2.

The capacitors in the phase shifter, plus others inserted in the feedback branches of the converter, also serve to increase the impedance of the converter to low frequency signals and dial pulses. By this technique, the inclusion of the shunt converter across the lines does not seriously decrease signaling ranges of most types of signaling systems used in Bell System exchange circuits.

4.3.5 *Comparison of Physical Converter Circuit to Ideal*

In the preceding discussions on principles of operation, the series and shunt units were first treated separately and then combined in a bridged-

T arrangement. The biasing and balancing networks arrangements were, however, omitted for simplification of presentation. These omissions are included in the complete schematic of the E6 converter shown in Fig. 12. The double-compound transistor arrangements used in both the series and shunt converter units are also indicated.

The principal assumptions made in the theoretical circuit descriptions were that transistor alphas were equal to unity and that the direct current biasing circuits could be neglected. In the following discussion, the

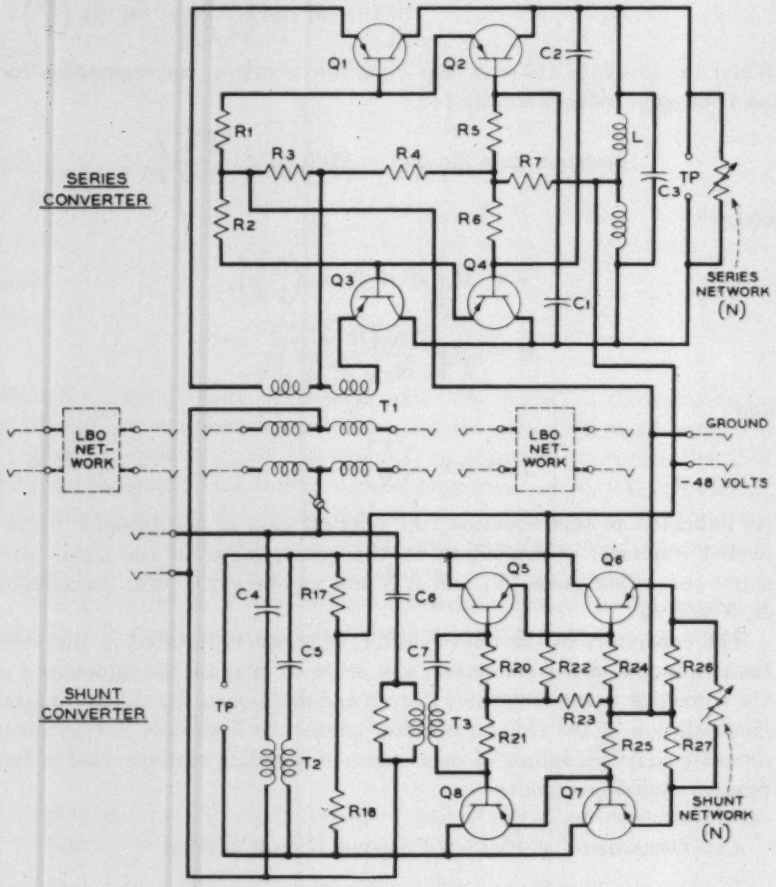


Fig. 12 — Complete schematic of E6 repeater.

validity of these assumptions will be assessed and the effects of other circuit elements on converter operation will be evaluated.

The double-compound transistor arrangement due to Darlington[6] has the advantage that the alpha of the compound transistor approaches unity more closely than the alpha of a single transistor. In addition, the use of such an arrangement insures that the over-all gain of the circuit is kept substantially constant regardless of normal variations in the individual transistor amplification factors. Each compound transistor, nevertheless, functions substantially as a single transistor and can be treated as such.

A more detailed sketch of the compound arrangement, corresponding to a single transistor, is shown in Fig. 13. Here the currents in the emitter, base and collector of each component transistor are defined. The compound alpha may be expressed as

$$\alpha_c = 1 - (1 - \alpha_1)(1 - \alpha_2).$$

Thus, if both transistors have a minimum alpha of 0.94, the compound alpha is found to be $\alpha_c = 0.9964$.

The DC bias arrangements of both the series and shunt elements of the converter are such as to have little effect on the angle of the conversion factor. Any changes in the real part of the conversion factor are small and can be compensated for in the series and shunt gain-controlling networks.

In the design of the E6 converters, the transformer T1 of Fig. 12 is considered as part of the series element of the converter. The purpose of this transformer is to combine the series and shunt elements of the converter in a bridged-T arrangement. The design of the transformer is quite critical in that the magnitudes of the leakage and mutual inductances and of the distributed capacitance parameters limit the maximum gain available from the series converter. The transformer design used

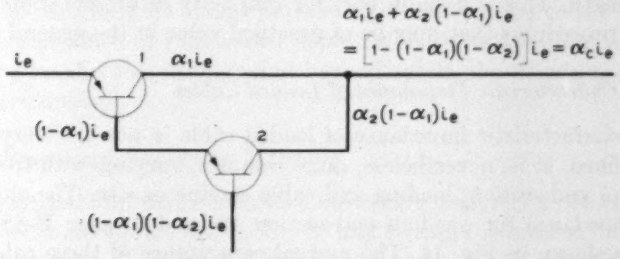


Fig. 13 — Compound transistor arrangement corresponding to single transistor.

in the E6 repeater, however, is not limiting up to gains of 13 or 14 db. It must, however, be capable of carrying the line currents due to talking battery or to signaling.

In the shunt-connected converter transformers T2 in the phase-shift network and T3 in the feedback circuit are used to isolate the DC bias arrangement from the effects of power-induced longitudinal voltages. In a small percentage of cables in the telephone plant, these 60-cps longitudinal voltages may be as high as 50 volts. The high mutual inductances and low series resistances of these transformers cause the transformers to have negligible effect in the transmitted band of the repeater. The use of these transformers also eliminates the longitudinal balance problem in the shunt connected converter. If the transformers were not specified, it would be necessary to introduce highly balanced capacitors in the tip and ring portions of the phase shifter and feedback circuits.

4.4 *Line Building-Out (LBO) Networks*

4.4.1 *General*

In a previous section it was pointed out that the impedance-matching properties of LBO networks were defined once the input impedance of the associated converter had been specified. Since the input impedance of the E6 converter is nominally 900 ohms in series with 2.16 microfarads (900 + 2.16 microfarads), LBO networks are required to match the characteristic impedance of variable end-section loaded cable facilities (of particular gage and type of loading) to this impedance.

A somewhat similar problem, which has been of interest to network designers for a number of years, is that of the broadband matching of an arbitrary impedance to a pure resistance. Although this general problem will not be considered in detail, the problem of matching the characteristic impedance of loaded cables to a pure resistance will be treated at some length. This is done in part for simplicity of presentation and to outline procedures that may be of practical value in the general case.

4.4.2 *Characteristic Impedance of Loaded Cables*

The characteristic impedance of loaded cable is not arbitrary, but is well-defined. It is, nevertheless, quite complex, varying with frequency, length of end-section, loading and cable conductor size. The characteristic impedance for one-half end-section 19- and 22-gage H-88 loaded cable is shown in Fig. 14. The mutual capacitance of these cable pairs is 0.082 microfarad per mile, and the frequency of cutoff is about 3480 cps.

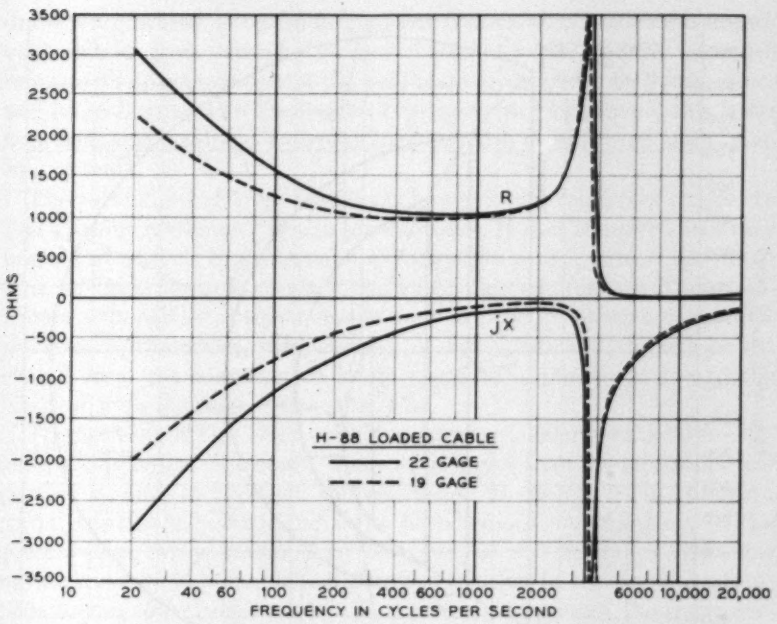


Fig. 14 — Characteristic impedance of 0.5 end-section H-88 loaded cable.

Not only does the impedance of these cable facilities vary with conductor size, but it varies with the length of end-section as well. The variation in impedance of 24-gage H-88 with varying length of end-section is shown in Fig. 15. The impedance characteristic below cutoff is indicated as the length of end section is increased from zero (adjacent to a loading coil) to full section (6000 feet from a loading coil).

The large variations in impedance shown in Figs. 14 and 15 complicate the job of matching the image impedance of the repeater to the characteristic impedance of the loaded cable. When it is specified that the matching be effective through and beyond cutoff, the situation is made even more difficult.

4.4.3 *Broadband Impedance Matching*

A broadband impedance-matching network is generally a two-terminal pair network similar to that of Fig. 16 and having the property that the input impedance Z_{11} approximates a pure resistance R when the network is terminated in an arbitrary impedance Z_x. For the realization of such a network, several design methods are available; however, an

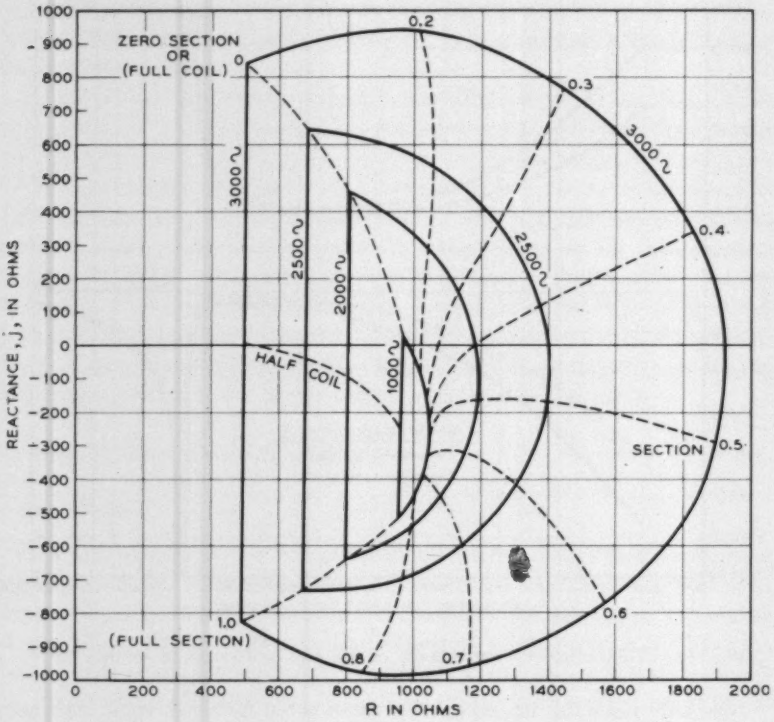


Fig. 15 — Impedances at various frequencies and variable end-sections for 24-gage H-88 loaded cable.

optimum network for a particular application depends on the impedance to be matched, the maximum tolerable flat loss and other requirements.

The case of lossless matching presents a continuing challenge. Basic limitations in terms of bandwidth and reflection coefficient are given by Fano,[7] but the technique is limited to a few special types of arbitrary impedance and is tedious to calculate. The design technique presented by Fano has been modified by Matthaei[8] to give straightforward proce-

Fig. 16 — General network arrangement for matching an arbitrary impedance to a pure resistance.

dures for synthesis of optimum, lossless, Tchebycheff, wideband impedance-matching networks for various classes of loads. Another approach requires an approximation of the conjugate of the given load impedance and its realization as a Darlington-type network.[9] In general, this latter approach also results in networks that are both complicated and expensive to build.

The techniques of lossy matching, however, are relatively simple. In Fig. 17 three methods of design are indicated. If Z_x is the given arbitrary impedance and Z_a is a rational approximation to Z_x, then it is evident that all three networks satisfy the matching requirement. These networks, although deceptively simple, may in practice become quite complicated. Furthermore, the networks of Fig. 17(a) and (b) will be excluded from consideration in many practical applications due to the inherent 6-db flat loss of the networks.

The network of Fig. 17(c) would also appear to offer considerable difficulties in practice. Obviously, R has to be larger than the maximum real part of Z_a and, in addition, the realization of the $(R - Z_a)$ function could reasonably be expected to be both tedious and complex. In the general case, this is undoubtedly true, but, as will be shown later, this network results in a simple and effective solution to the broadband impedance matching problem when Z_x is taken to represent the characteristic impedance of loaded cable facilities.[10]

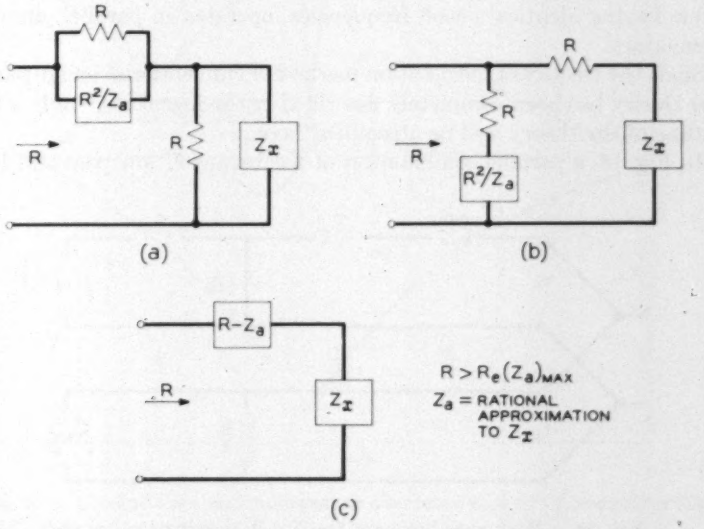


Fig. 17 — Lossy networks matching an arbitrary impedance to a pure resistance.

4.4.4 *Practical LBO Networks*

In the development of the E6 repeater, requirements placed on the LBO network precluded the use of any of the previously described methods of approach with the single exception of the method indicated in Fig. 17(c). Requirements on return loss, echo return loss and insertion loss were severe, but the added requirements on allowable size and cost were limiting, in that complex networks were thereby excluded. Requirements on low-frequency signaling margins were particularly severe, and these effectively established criteria for network configurations that might be employed. Other requirements relating to longitudinal unbalance, protection against high voltage surges, etc., also had to be considered.

The application of immitance-matching techniques to the design of LBO networks, however, has resulted in a number of practical networks. Three different approaches considered during the development are significant. These are: (a) the fractional termination method of image parameter theory; (b) the "lossy" matching method implied in Fig. 17(c) and (c) the method of impedance compensation.

4.4.4.1 *Fractional Termination Method.* This method of design is based on recognition of two essential facts: (a) a midsection loaded cable can be simulated approximately by a midshunt constant-K image impedance and (b) properly terminated constant-K low-pass and high-pass filters having identical cutoff frequencies, operated in parallel, are complementary.

Since the fractional termination method of conventional image parameter theory has been completely described in the literature,[11] only a brief outline of the theory will be attempted here.

In Fig. 18, a parallel combination of a constant K low-pass and high-

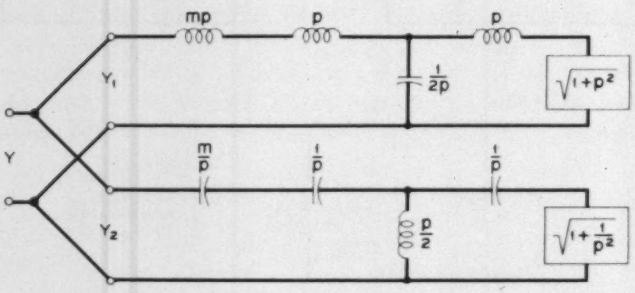


Fig. 18 — High-pass–low-pass fractional termination network.

pass filter is shown. The impedances and frequencies are normalized to unity. Thus,

$$Y_1 = \frac{1}{mp + 1 + p^2},$$

$$Y_2 = \frac{p}{m + 1 + p^2}.$$

In the low-pass band, $p = j\omega$, $\omega \leqq 1$; hence,

$$Y_1 = \frac{1 - \omega^2}{m^2\omega^2 + 1 - \omega^2} - \frac{jm\omega}{m^2\omega^2 + 1 - \omega^2},$$

$$Y_2 = \frac{j\omega}{m + 1 - \omega^2}.$$

Similar expressions for the high-pass band can be found. The real and imaginary parts of Y_1 and Y_2 are plotted in Fig. 19. It is recognized that

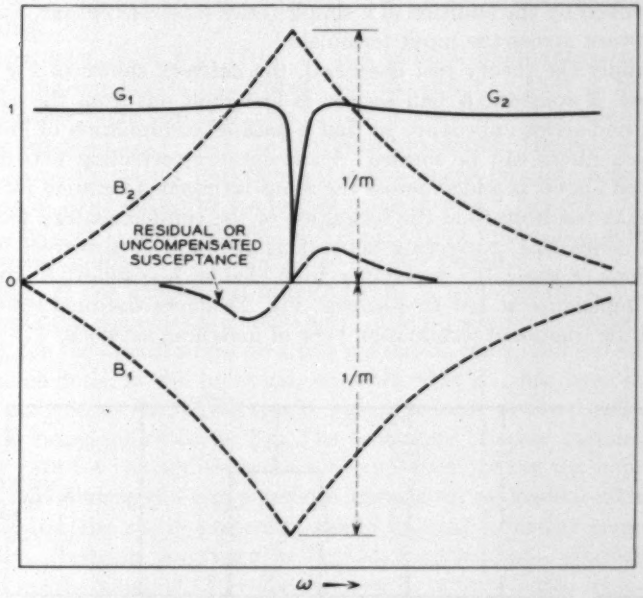


Fig. 19 — Conductance and susceptance characteristics of fractional termination network.

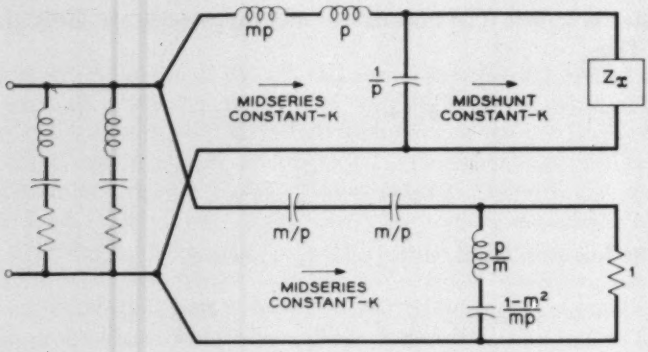


Fig. 20 — Fractional termination network with image impedances replaced by physically realizable terminations.

the real parts are the m-derived image impedances; hence, the input admittance Y is, roughly, a constant resistance except for residual effects in the vicinity of cutoff. The matching properties of the network may be improved by the addition of a simple series RLC susceptance-correcting network across the input terminals.

To apply the theory just described, the network shown in Fig. 20 is designed. A constant-K half section is first built out from the cable to form a mid-series impedance so that a parallel combination of low- and high-pass filters can be formed. A susceptance-correcting network, as described above, is added across the input terminals to correct for residual effects resulting from the formation of the complementary filters. A second series RLC correcting network is also added to correct for the deviations of the cable impedance from that of a constant-K low-pass image impedance at low frequencies. Fig. 21 shows the final configuration of the fractional termination type of matching network.

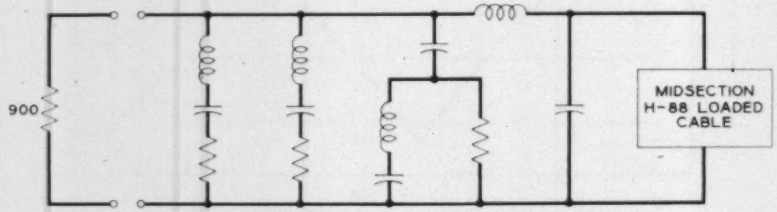


Fig. 21 — Two-terminal pair fractional termination LBO network.

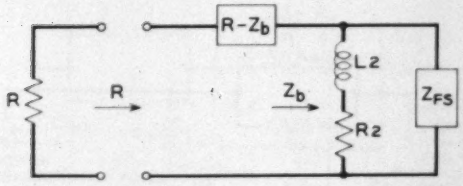


Fig. 22 — Lossy matching of full section of loaded cable to a pure resistance.

4.4.4.2 *Lossy Matching.* Although the fractional termination method of design results in practical networks, these networks are too complex from a cost point of view. In addition, the networks are limited to the extent that the cable impedance can be built out to represent a midshunt constant-K image impedance.

One of the simplest approaches to the design of LBO networks is based on the realization of a network that, when placed in series with the cable impedance, results in a pure resistance match. Such a network, as shown in Fig. 17(c), was described in the general discussion of lossy matching.

In the network of Fig. 17(c), Z_a is a rational approximation to the terminating impedance Z_x. Except for the criterion that R be larger than the maximum real of Z_a, no restrictions are placed on the approximating function $(R - Z_a)$. As previously indicated, the realization of such an approximating impedance function might be quite difficult. In the case of LBO networks for loaded cable facilities, however, the synthesis problem is quite simple, provided certain impedance correcting techniques applicable to loaded cable are recognized.

In Fig. 22 the generalized impedance Z_x is replaced by Z_{FS}, which represents the characteristic impedance of a full section of loaded cable. Except for the specification of a full section of cable, this substitution is quite obvious. It will be noted, however, that Z_{FS} has been shunted by a simple series RL network, with the impedance of the parallel combination being indicated as Z_b. The technique of lossy matching obviously calls for the addition of a series network having the impedance $(R - Z_b)$. Unless the characteristic impedance of loaded cable were investigated, the above procedure would indicate a further complexity in design. However, as shown in Fig. 23, a considerable simplification results, for the function $(R - Z_b)$ can now be approximated by a simple series RC network in parallel with an inductor.

Assuming that the simple network described above is capable of yield-

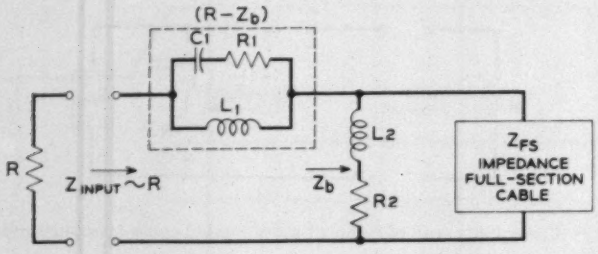


Fig. 23 — Lossy matching approximation to synthesis of LBO networks.

ing satisfactory results, it would appear that certain difficulties still exist. The network requires a full-section cable termination. Fortunately, the impedance of variable end-sections of cable can be easily built out to full-section value by the addition of series resistances and shunt capacitances.

Although the network of Fig. 23 is capable of meeting most of the requirements specified, it cannot be used without modification because of the signaling penalities that result from its application. The series RL network connected across the full-section cable impedance shunts direct currents, presents a low impedance to signal frequencies and degrades service. By placing a sufficiently small capacitor in series with the series RL network, the signaling penalties can be reduced to any desired extent, but the effectiveness of the admittance correcting network R_2L_2 is correspondingly impaired.

A compromise must, therefore, be reached. For signaling purposes it is possible to specify a minimum value of capacitance that can be tolerated. Once this has been done, a new network may be designed following the procedures previously outlined. Since the resulting network represents a compromise, degradation in certain characteristics must be expected. The complete network including the building-out capacitors, building-out resistors and the "signaling" capacitor is shown in Fig. 24.

4.4.4.3 *Impedance Compensation Method.* In order to describe the impedance compensation method adequately, it is necessary to reconsider the variations that take place in cable impedance as the length of end-section is varied. In this instance, it is more convenient to plot the real and imaginary components of the cable impedance as functions of frequency.

The wide variations that take place in the characteristic impedance of loaded cable as the length of end-section is varied is indicated by the solid curves of Figs. 25, 26 and 27 for 0.2, 0.5 and 1.0 end-section. It

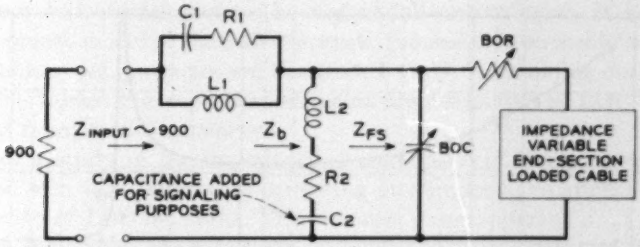


Fig. 24 — Practical LBO network for matching variable-end-section loaded cable to 900 ohms.

will be observed that the resonance effects in the vicinity of cutoff either disappear or become highly damped as the length of end-section is increased.

As noted previously, variable end-sections of cable may be built out to full-section value by the addition of series resistance and shunt capacitance. The effect of adding only capacitance to the fractional end-section is indicated by the same figures. It may be seen from comparisons between the fractional and full end-section characteristics that a series resistance must be supplied to raise the resistive component of a cable

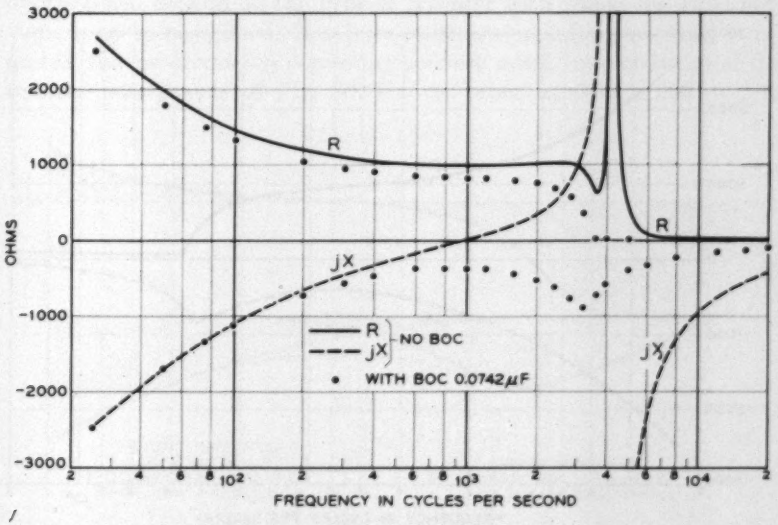


Fig. 25 — Characteristic impedance of 0.2 end-section 22-gage H-88 loaded cable with and without building-out capacitance network added.

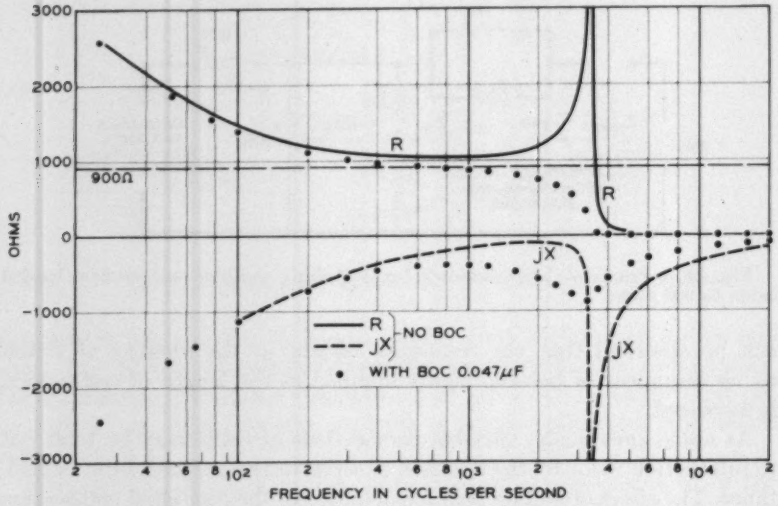


Fig. 26 — Characteristic impedance of 0.5 end-section 22-gage H-88 loaded cable with and without building-out capacitance network added.

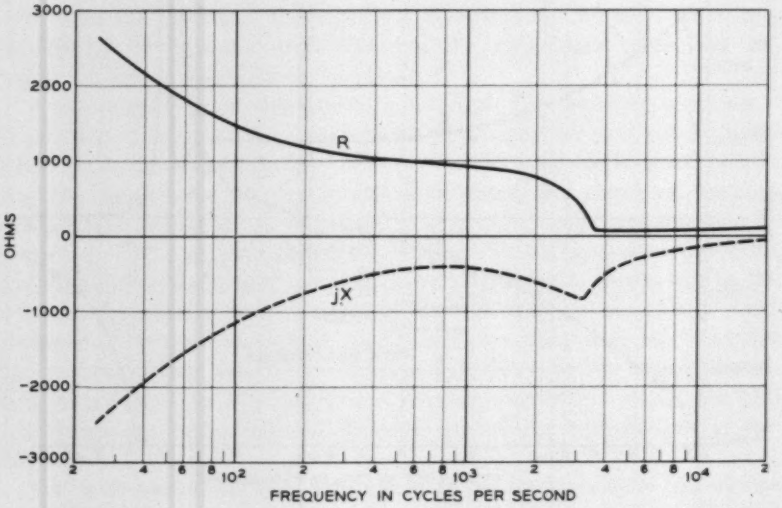


Fig. 27 — Characteristic impedance of full-section H-88 loaded cable.

built out with capacitance to the desired full-section value. In the final development model of the LBO network, the resistive elements supplied for building out purposes are designated as BOR (building-out resistance) networks. The capacitive elements are designated as BOC (building-out capacitance) networks.

From a study of the impedance characteristic of built-out cable, it may be seen that impedance-correcting techniques have application to the design of LBO networks. This becomes quite apparent if, as in Fig. 28, the impedance characteristic is divided into two zones of interest along the frequency axis. The first of these zones, the high-frequency correcting (HFC) zone, requires a series impedance type of correction. The second, the low-frequency correction (LFC) zone, requires a shunt impedance or admittance type of correction.

The problem of impedance correction in the high-frequency zone is easily solved by application of the simple network of Fig. 29(a). The schematic shows both the network and the general shape of its two-terminal impedance characteristic. In Fig. 29(b) the corrected impedance of the built-out cable is indicated for the idealized case. If the admittance characteristic of the high-frequency corrected cable impedance is now plotted, a characteristic similar to that of Fig. 30 results.

It is quite obvious that the design of the low-frequency correcting network will depend on the type of repeater with which the LBO network is to be associated. If a pure-resistance matching network is required, the low-frequency correcting network must completely annul the residual susceptance of Fig. 30. On the other hand, if a 900 + 2.16

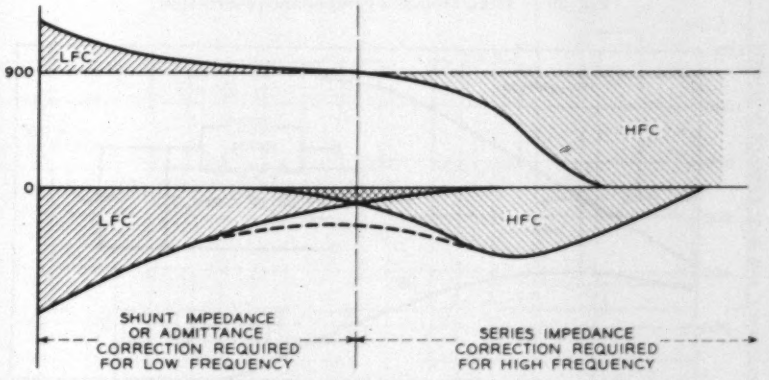


Fig. 28 — Division of LBO network design into low- and high-frequency correction zones.

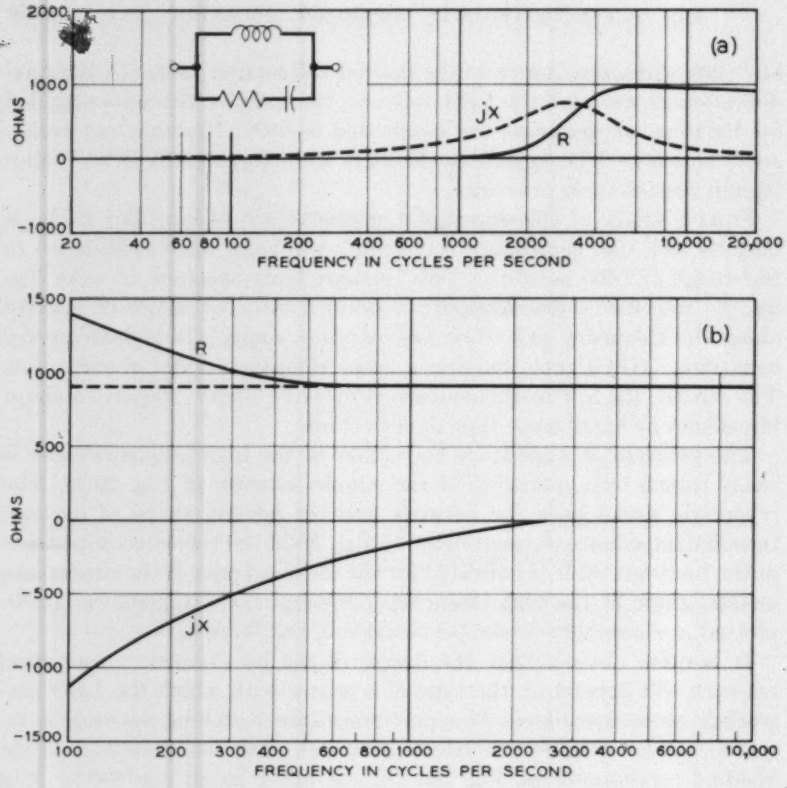


Fig. 29 — High-frequency impedance correction.

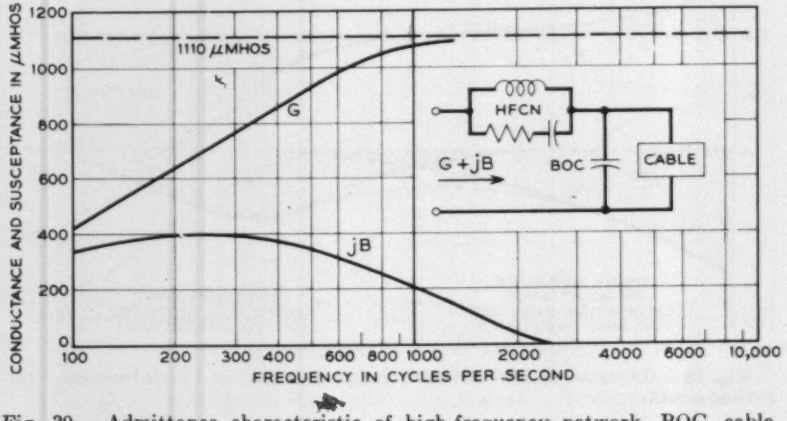


Fig. 30 — Admittance characteristic of high-frequency network, BOC, cable.

microfarad network is required, the susceptance need be annulled only partially. As previously noted, a series capacitor must be included for signaling and supervisory reasons.

4.4.5 *Description of Final Development of LBO Network*

In previous sections, the LBO networks were shown as unbalanced structures. Inasmuch as these networks must work with balanced cable pairs, however, the actual network requires balanced construction for the BOR and HFC networks.

As indicated in Fig. 31, the final development model of the LBO network (designated as the 830-type network) includes a resistance (BOR) and a capacitance (BOC) network for building out cable pairs to full-end section value. These networks are required for all gages of cable. The high-frequency correcting network represents a compromise in design suitable for all gages of cable. The three-winding transformer indicated in the sketch inserts the required balanced inductance in each series branch of the network and effectively places the single *RC* network termination of the third winding in shunt with each inductance. Problems of longitudinal balance are, therefore, minimized and only one *RC* network is required.

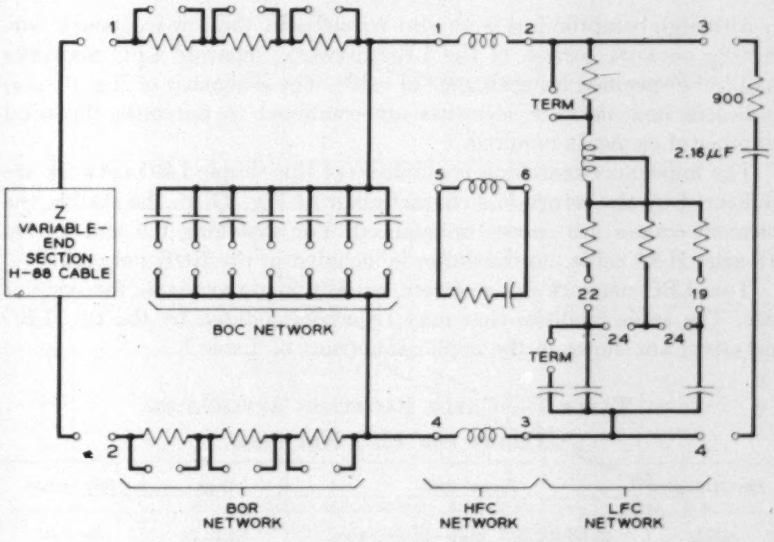


Fig. 31 — Complete line building-out network for H-88 cable facilities.

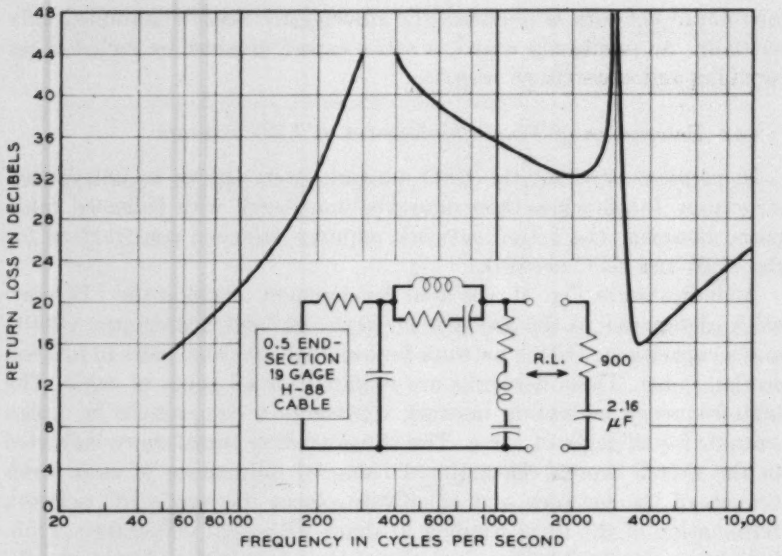


Fig. 32 — Return-loss characteristic of line-building-out network for 0.5 end-section 19-gage H-88 cable.

Although compromises were also required in the low-frequency correcting network section of the LBO network, separate LFC networks had to be specified for each gage of cable. The schematic of Fig. 31 also indicates how network elements are combined to minimize the total number of elements required.

The impedance-matching possibilities of this simple LBO network are indicated by the return-loss characteristic of Fig. 32. In the sketch, the element values are shown unbalanced. For matching 0.5 end-section 19-gage H-88 cable, no resistance is included in the BOR network.

Two LBO network designs were initially made available for systems use. The cable facilities that may be accommodated by the two LBO networks are shown in the application chart of Table I.

TABLE I — CABLE FACILITIES APPLICATION CHART FOR LBO NETWORK

LBO Network	Type of Cable	Gage	End-Section
830A	H-88 loaded, high capacitance	19,22,24	0—full
830B	H-88 loaded, low capacitance	19	0—full
830B	D-88 loaded, high capacitance	22	0—full

V. PERFORMANCE CHARACTERISTICS OF THE E6 REPEATER

5.1 *Impedance and Return Loss*

A typical example of the impedances existing at the converter-LBO network junction is shown in Fig. 33. The dashed lines show the impedance of 24-gage H-88 cable as modified by the LBO, and the solid lines show the impedance of a gain unit having a gain of 6 db. Curves for other cable facilities and other gains are very similar, because the combined converter impedance is quite constant with gain while the LBO networks correct for the differences between the several cable facilities.

A method for checking that the series and shunt converters have the correct relative impedance for any given gain setting is to make the individual 1000-cycle insertion gains of these converter units the same when they are each measured between image impedances of the converter. If this adjustment procedure is correct, the image impedance of the converter should result when the expression for the loss of the series-connected converter is equated to that of the shunt-connected converter.

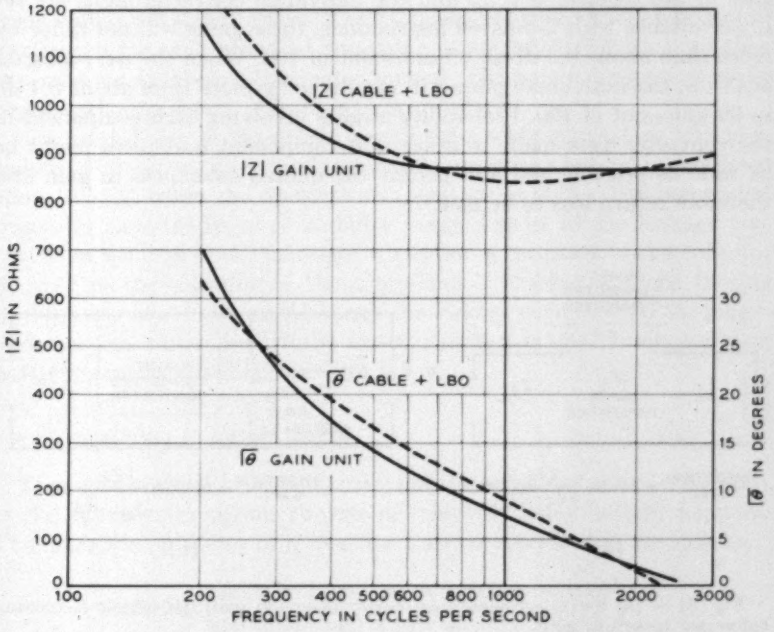


Fig. 33 — E6 gain unit impedance compared to impedance of 24-gage H-88 cable plus LBO.

In Fig. 34, a circuit arrangement used for individual series and shunt unit impedance adjustment by measurement of insertion gain is shown. Equating the two expressions indicated in Fig. 34,

$$20 \log_{10} \frac{2Z_I}{2Z_I - Z_A} = 20 \log_{10} \frac{-Z_B}{-Z_B + \frac{Z_I}{2}}.$$

Solving for Z_I, the familiar expression for the image impedance of the bridged-T negative impedance converter results:

$$Z_I = \sqrt{Z_A Z_B}.$$

The return loss between the E6 converter and the LBO and cable is about 34 db between 500 and 2500 cycles when the individual converter insertion gains are alike and the cable has nominal impedance. When the individual converter insertion gains differ by 0.2 db, this return loss is about 32 db, and when they differ by 0.4 db, it is about 28 db.

Tolerances on converter components are such that, when the over-all gain of the repeater is 6 db and the individual converter gains are set in accordance with tabulated instructions, these gains will not differ by more than about 0.2 db in 99 units out of 100. When the over-all gain is 13 db, the individual gains will not differ by more than about 0.4 db in 99 units out of 100. Probability studies involving each component in the converter were made in order that component tolerances might be as wide as possible and still permit the quoted tolerances in gain and therefore return loss to be met.

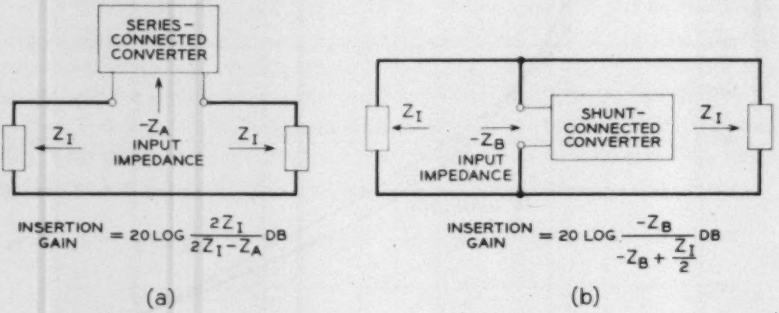


Fig. 34 — (a) Series-connected converter insertion gain; (b) shunt-connected converter insertion gain.

5.2 *Load Capacity and DC Stability*

The maximum output level that a repeater should be designed to carry is determined by cable crosstalk, allowable repeater gain and maximum talker volumes. From these points of view, it was found desirable that the E6 repeater carry approximately a +16 dbm single-frequency output level with relatively little compression.

In the E6 repeater the currents from the shunt and series converters add in phase in the receiving line and cancel in the sending line. Thus, all of the power from the converters is available at the receiving line. However, the maximum power available there decreases as the converter gain is increased. This is mainly because the gain-adjusting network of the series converter absorbs more power as the gain-adjusting network is increased to give more gain.

The maximum available power from the transistors is a function of the ambient temperature and their ability to dissipate heat due to the DC biases. The transistors used in the E6 repeater will stand 0.150 watt at 50°C and 0.110 watt at 60°C. Maximum ambient temperature is considered to be 50°C in most cases, but 60°C may be encountered in some unmanned desert offices. The DC biases are such that the DC power dissipation for each principal transistor is roughly 0.085 watt, so they are operated well below their rated output. The power dissipation in the piggy-back transistors is from 4 to 8 milliwatts.

The DC stability of the circuit is such that, when transistors with high leakage (i_{c0}) currents are used, the maximum power output is decreased about 0.5 db when the ambient temperature is 60°C. The compound transistor circuitry helps DC stability because most of the leakage current from the base of the principal transistor is returned to the collector by way of the collector of the piggy-back transistor without flowing through any bias resistors. Although the leakage current from the piggy-back transistor does disturb the biases, its effect is small because the DC power dissipation in this transistor is low.

5.3 *Susceptibility to Battery Noise*

The well-balanced push-pull circuits of the series and shunt converters of the E6 repeater permit the use of relatively noisy battery supplies. The noise susceptibility of a repeater may be expressed in decibels as

$$20 \log \left(\frac{\text{noise voltage across repeater output}}{\text{noise voltage across battery supply}} \right).$$

The noise in the output of the E6 repeater is at least 53 db lower than that in the battery supply at all frequencies between 60 to 3000 cps. Thus, a fairly noisy battery with a noise voltage of +45 db RN, as read on a Western Electric 2B noise set, would cause a repeater output noise of about (+45 − 53) db or −8 db RN.

5.4 *Lightning Surge Protection*

Lightning surge voltages are limited by the cable main frame protector blocks to about 600 volts peak. The average decay of surges is such that this peak voltage is decreased to about one-half in 600 microseconds. In some offices, depending on exposure and the number of storms, there may be as many as several hundred hits, with peak values from 200 to 600 volts, occurring in a single year. The total number of hits on a single cable pair during the lifetime of an E6 repeater may number in the thousands.

The circuits of the E6 converters are such that when 600-volt longitudinal or transverse surge is applied to its line terminals the maximum emitter-to-base (or collector-to-base) current through the transistors is about 0.5 ampere. The duration of this pulse is of the same order as that of the surge voltage (i.e., one-half peak value in 600 microseconds). The transistors used in the E6 repeater are designed to accept such surges without damage.

VI. APPARATUS DESCRIPTION OF THE E6 REPEATER

The mechanical design of the E6 repeater, a photograph of which is shown in Fig. 35, departs from conventional methods of packaging and is based on the matrix structure. In the E6 repeater, therefore, the printed wiring boards are used only for interconnection of the components into the desired electrical configuration as shown in Figs. 36 and 37. Molded phenolic blocks, sandwiched between the printed wiring boards, are used for mechanical support. Shaped cavities, within the molded blocks, are designed to accept the individual components. Two types of cavity are available within the blocks: one for pig-tail components and another for the larger components utilizing flexible leads.

The pig-tail cavities are so arranged that the axial dimensions of assembled component apparatus are at right angles to the printed wiring boards. To facilitate assembly, that portion of the block designed to accept pig-tail components is split into two equal and mating parts. The components are inserted into the lower half of the split block, which

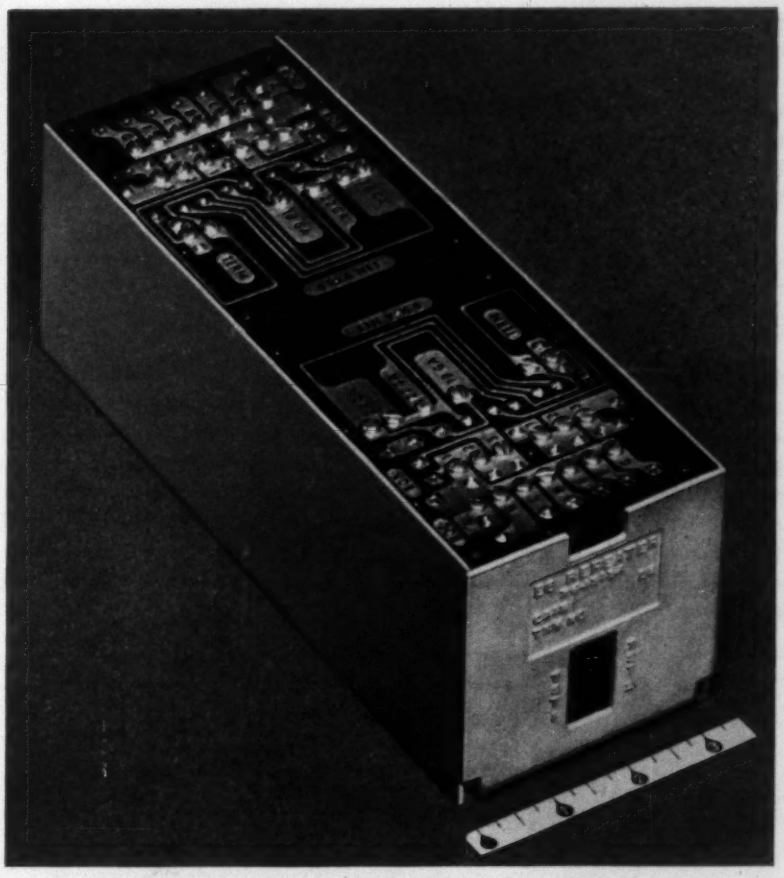


Fig. 35 — The E6 repeater.

acts as a carrier during manufacture. The top half is then dropped over the lower portion of the block in final assembly.

The portion of the phenolic block reserved for flexible lead components is not split, but is designed to supply open-top cavities into which components may be placed and subsequently potted. Connection to the printed wiring boards is made by means of feed-through terminals to which the flexible terminal leads of the components are soldered prior to encasement in a microcrystalline wax. In the converter the open-top

cavities are protected by a metal plate, while in the LBO network this function is performed by one of the printed wiring boards.

To facilitate the insertion of pig-tail components, the individual cavities are tapered or funnel-shaped, with an additional tapered section at the base to guide the pig-tail lead wire through to the printed wiring board. The pig-tail leads are precut to proper length for connection to the printed conductors.

Transistors present the only exception to the cavity method of mounting, primarily because of the need for heat dissipation. A shelf was therefore provided on the lower block of the converter unit for external assembly of the transistors.

The converter unit of the E6 repeater, illustrating the features described above, is shown in Fig. 36. The lower portion of the phenolic mold is shown at the left of the photograph. The individual pig-tail components have been assembled in their individual cavities and the top portion of the pig-tail section is about to be placed in position. The open cavity section of the mold, reserved for flexible lead components, is also shown. In this photograph, the microcrystalline wax has not yet been poured into the open-cavity molded section. In normal production, however, this would have been done prior to the assembly of the pig-tail components.

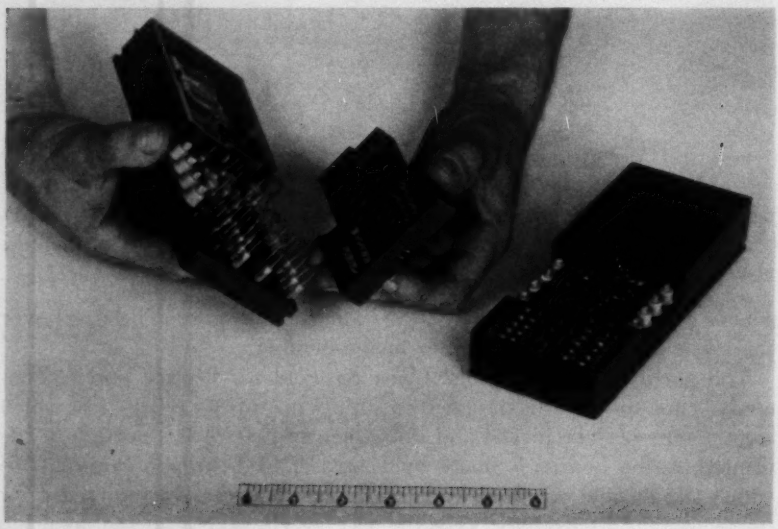


Fig. 36 — The E6 negative impedance converter.

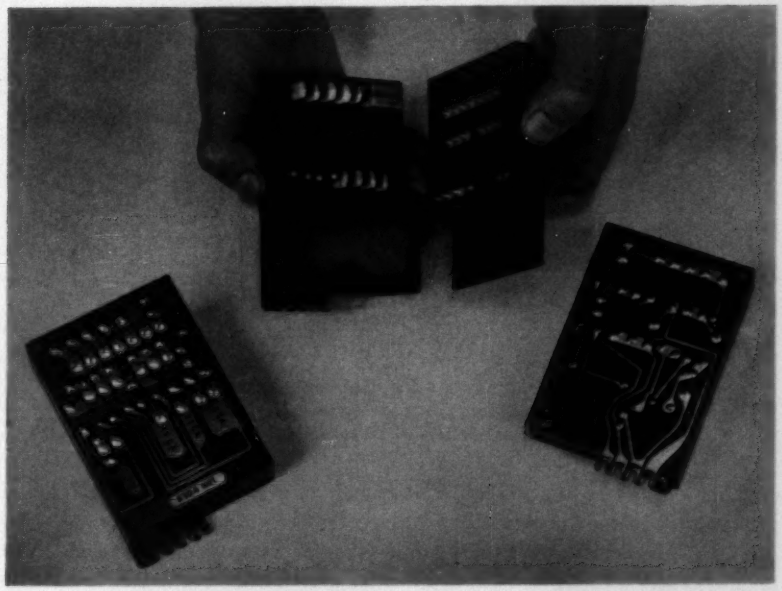


Fig. 37 — The 830A line building-out network.

Similar views of the LBO network are shown in Fig. 37. In this instance, the microcrystalline wax has already been poured and the unit is ready for assembly of the top section of the mold. In the other two views, the split-molded block is shown sandwiched between the top and bottom printed wiring boards of completely assembled LBO networks. The comb-like protrusion in the phenolic block, at one end of the LBO network, permits connection of the LBO networks to the converter. Electrical connection is made through open-ended brass grommets, soldered to the printed wiring board and arranged to accept the screws that go into tapped metal bushings located in the converter.

The E6 repeater is a plug-in unit, and all external connections are made by means of the plug and guidance details associated with the converter unit of the repeater, as shown at the right of Fig. 38. The mating connector is mounted on the equipment shelf. The connector assembly is such that a floating action at each terminal is obtained. When the repeater unit is plugged in on its mounting shelf, the tapered prongs of the guidance detail align the plug and the connector prior to

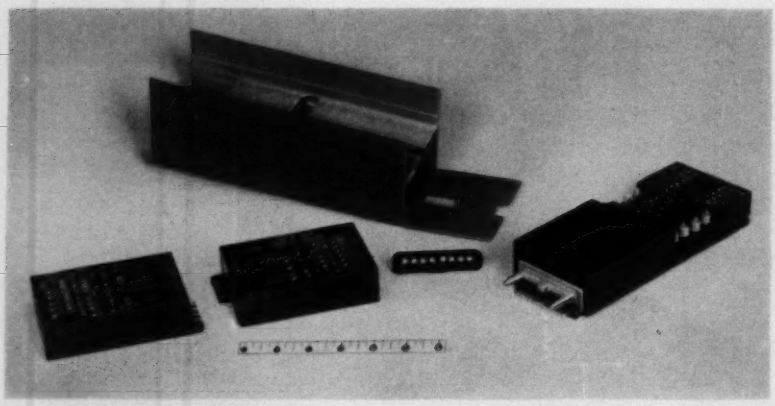


Fig. 38 — Exploded view of complete E6 repeater.

engagement. A spring catch on the shelf engages one of the guidance prongs and locks the repeater in place.

The E6 repeater consists of one converter unit and one or two LBO networks as dictated by systems use. As previously indicated, two LBO networks are required for intermediate use, while only one is required for terminal applications. The converter and LBO networks are housed in the aluminum H-beam extrusion of Fig. 38, which shows an exploded view of the complete E6 repeater. The bottom compartment of the H-beam extrusion accepts the converter unit, which is inserted guidance detail first from the front or hinged-door end. Grooves, molded in the sides of the converter unit, mesh with rails extruded along the inner surface of the H-beam. The transistors and adjusting screws, shown in the photograph are accessible at the bottom of the beam, when the converter is inserted in the housing.

The LBO networks are mounted in a similar fashion to that employed for the converter. The two networks, however, are each fed into the housing from opposite ends, one from the hinged-door end, and the other from the open end. The networks are positioned so that the comb-like protrusions of each network align in position above a filler block for screw connection to the converter.

For terminal applications, although only one LBO network is required, a dummy network must be supplied to provide a wiring path from one side of the converter to the line connections. In this instance, as shown in Fig. 39, the LBO network inserted from the hinged-door end is retained.

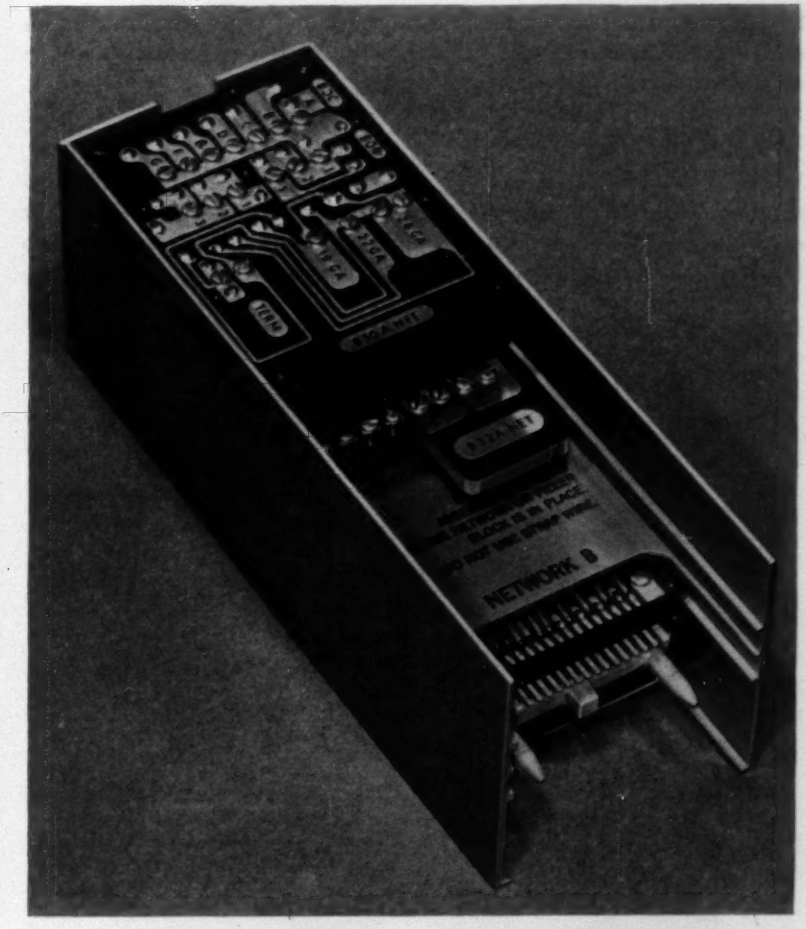


Fig. 39 — The E6 repeater — arranged for terminal use.

VII. TRANSMISSION TESTING EQUIPMENT

7.1 *Introduction*

One of the principal objectives of the E6 development was to reduce the cost of installing and maintaining negative impedance repeaters in the exchange area plant. In addition to the basic contributions of the E6 repeater toward this end, further simplification and economies have been

realized by making available three new pieces of test equipment. These are the 54A transmission measuring set, the 54B test stand, and the 54C return loss measuring set.

7.2 *54A Transmission Measuring Set*

The 54A test set, shown in Fig. 40, is a portable, transistorized transmission measuring set used to determine the gain of the series and shunt converters of the E6 repeater — individually, or in combination. The importance of this function is evident when it is recalled that the attain-

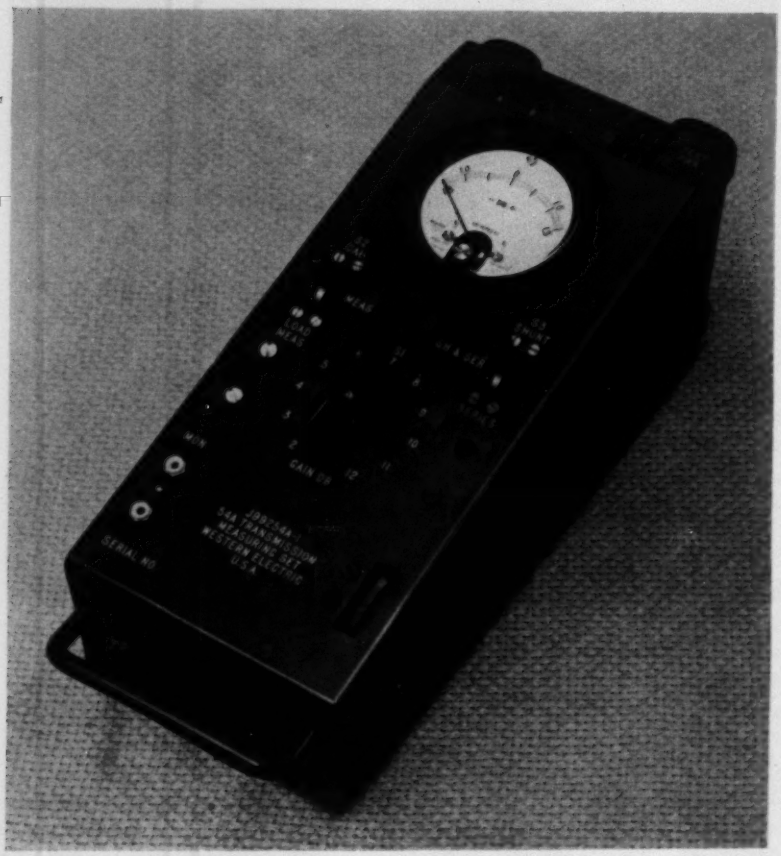


Fig. 40 — 54A transmission measuring set.

ment of high return loss between the converter and LBO networks of the E6 repeater is dependent on the equality of gain settings of the series and shunt converters. The set may also be used for periodic measurements of gain and signal-level carrying capacity of the repeater.

7.3 *54B Test Stand*

The E6 repeater cannot be adjusted in position on the relay rack mounting shelf. In this position, the adjusting screws controlling converter gain and LBO network impedance-matching functions are not accessible. The 54B test stand, shown in Fig. 41, is therefore provided. The purpose of the stand is: (a) to hold the repeater in a convenient position for adjustment of the gain units and the LBO networks; (b) to supply DC power connections for the 54A transmission measuring set, the 54C return loss measuring set and the E6 repeater; and (c) to provide convenient connections and points of access for gain, return loss and trouble-locating tests.

The test stand is provided with a rotating platform into which the E6 repeater is inserted. The platform has two positions, each 180° removed from the other. In one position, the gain-adjusting screws of the converter are presented for adjustment, while in the other position the BOR, BOC and LFC network screws of the LBO network are made available.

Another feature of the test stand is the inclusion of a shunt holding circuit to prevent dialed-up test terminations from being released during test.

7.4 *54C Return Loss Measuring Set*

In order to attain the high degree of performance of which the E6 repeater is capable it is necessary that the LBO network and connected cable impedance match that of the E6 converter. In many applications, adequate return losses can be obtained at the LBO–converter junction by "book-value" adjustment of the BOR, BOC and LFC networks of the LBO network, based on office records of length of end-section and type of cable. This assumes, however, that the end-section value of the cable pair terminating a particular LBO network is accurately known. Experience on several field trials indicates, however, that the cable records are often not sufficiently accurate for this type of specification. The 54C return loss measuring set was, therefore, developed to permit optimum adjustment of the LBO networks on the basis of return loss measurements.

Fig. 41 — 54B test stand.

The 54C set, shown in Fig. 42, measures return loss in two frequency bands — either 500 to 2500 cps or 2000 to 3000 cps — with a self-contained sweep oscillator. Sweep frequencies are used because the relation between return loss and frequency can be highly variable and unpredictable in shape for specific cases. Such a sweep-frequency measurement

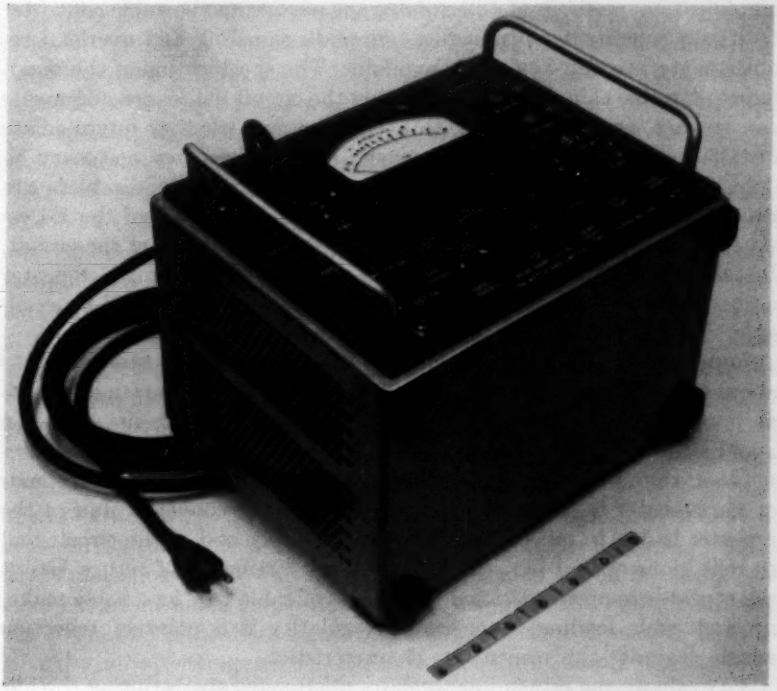


Fig. 42 — 54C return loss measuring set.

gives an average or integrated measurement that is more nearly representative than any arbitrary single-frequency measurement could be. The higher-frequency band is for BOC adjustment, which is more effective near cable cutoff; the lower and broader band for BOR and "gage" adjustments, which are effective across the whole band.

While the primary purpose of the 54C test set is for LBO network adjustment, it may ultimately be used in several general-purpose applications such as cable acceptance tests, adjustments of impedance compensators and precision networks associated with two-wire toll-connecting trunks.

VIII. ENGINEERING CONSIDERATIONS

In engineering trunk circuits for E6 repeater use, a number of factors must be considered if transmission loss objectives are to be met. Since the primary purpose of the E6 repeater is to provide voice-frequency

amplification, methods of determining the maximum allowable converter unit gain consistent with singing, crosstalk signaling and overload requirements, etc. must be made available. The specification of the maximum allowable gain determines whether the circuit net loss requirements can be met, and will dictate, to a large extent, whether intermediate installation of repeaters is necessary. It then becomes necessary to determine if idle circuit terminations or repeater gain disablers are needed to prevent singing during idle periods. The effect of the E6 repeater on signaling must also be considered in the design of the circuit. Installation instruction must then be provided for an initial repeater adjustment so that a starting point for line-up procedures will be available.

Repeaters located at the terminal end of trunk circuits are limited to about 6.5 db net gain due to overload and crosstalk considerations. Those at intermediate locations may be operated up to 12 db net gain. To avoid singing in the idle circuit condition and to provide a satisfactory net loss, a computation of probable return loss at the converter terminals of the repeater is required. The return loss at the converter unit of the repeater is due to reflections at the central office end of the circuit and to reflections caused by cable irregularities. Estimates of return loss at the repeater require a detailed knowledge of cable pair loss, cable makeup and cable loading, since each irregularity is a point of reflection which degrades the return loss characteristic.

To evaluate the return loss at the converter terminals during the idle circuit condition, either of two approaches may be taken. In the first case, the far-end terminals are considered to be open circuit without idle circuit terminations. The terminal return loss is assumed to be 0 db and the maximum allowable gain is computed. This usually results in a circuit net loss greater than 3 db effective.* In the second case, where lower net losses are required, one practice is to terminate the circuit during the idle condition in a resistance at the far end. An alternative approach for this case is to provide a repeater disabler to reduce repeater gain when no supervisory current flows in the line, that is, when the line is idle and therefore unterminated. Under either condition, the repeater gain can be raised to the point where singing occurs in the interval between seizure of the circuit and the establishment of a through connection to the distant end. It is customary to assume 5 db return loss at a point equipped with an idle circuit termination. This usually permits raising the repeater gain by 1 or 2 db.

* Average of 15 selected frequencies throughout the voice band.

Having determined the maximum allowable gain, the circuit net loss must be computed. The net loss is the sum of the office, line and LBO network losses less the gain of the converter. The LBO loss, however, depends on the length and make-up of the end-sections of cable facing the repeater, which determine the amount of building-out capacitance and building-out resistance required in the LBO network.

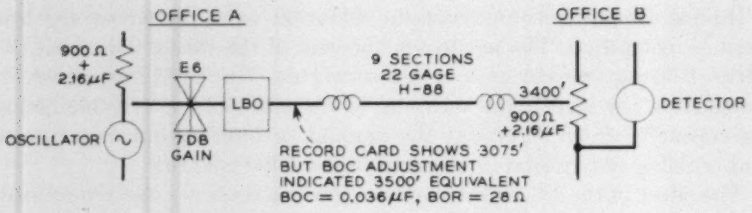
The effect of the E6 repeater on DC signaling must also be determined. The DC resistance of the converter and LBO networks introduces an adverse effect on signaling. The total shunt capacitance of the repeater introduces another. Some circuits, unfortunately, permit little margin for possible degradation of signaling due to repeater operation. In some instances, adjustment of signaling circuit relays may provide additional margins, but in others it may be found desirable to omit or reduce the BOR adjustment. This results in a degradation in return loss and thereby lowers the permissible gain.

Full realization of the return-loss capabilities of the E6 repeater can be obtained only if LBO network adjustments are made with a high degree of accuracy. The greatest problem encountered here is that of determining the value of end-section of a particular cable facility. Exact determination of cable end-section is complicated by office wiring and the mixture of cable gages that usually comprise the end-section. By use of the 54C return loss measuring set, however, it is possible to optimize the initial settings of the LBO network and to operate the converter at the maximum gain that cable irregularities will permit.

IX. SYSTEM RESULTS AND COMPARISON WITH E23 REPEATER

The E6 repeater will, in practically all cases, out-perform the older E23 repeater. With reasonable cable structural return losses, trunks equipped with E6 repeaters can be made to meet the systems requirements for over-all net loss and for return loss to the toll office balancing network. Much of the improved performance of the E6 repeater, as compared to the E23 repeater, is due to the range and precision of adjustment available in the LBO networks. Another contributing factor to improved performance is that the image impedance of the converter approximates 900 ohms + 2.16 microfarads and is essentially constant with gain setting.

Trunks equipped with E6 repeaters can usually be operated at 1 to 2 db lower net loss than can trunks with E23 repeaters. This lower loss is obtainable with equal or improved return losses at the ends of the trunk. In addition to the better performance, it is possible, by use of careful adjustment procedures, to raise the gain of E6 repeaters by a



MEASURED NET LOSS

	1000 CYCLE		EFFECTIVE	
	WITHOUT REPEATER	WITH REPEATER	WITHOUT REPEATER	WITH REPEATER
LAB SETUP	8.3 DB	1.7 DB	8.6 DB	2.3 DB
ACTUAL	9.0 DB	2.3 DB	9.2 DB	2.9 DB

RETURN LOSS AT OFFICE A WITH COMPENSATOR NETWORK TERMINATION AT OFFICE B

	500–2500 ∿	2000–3000 ∿	MINIMUM	CABLE STRUCTURAL RETURN LOSS
ACTUAL	20.3 DB	14.4 DB	8.3 DB AT 2900 ∿	38.1 DB
LAB SETUP	20.6 DB	19.0 DB	17.0 DB AT 2900 ∿	> 40.0 DB

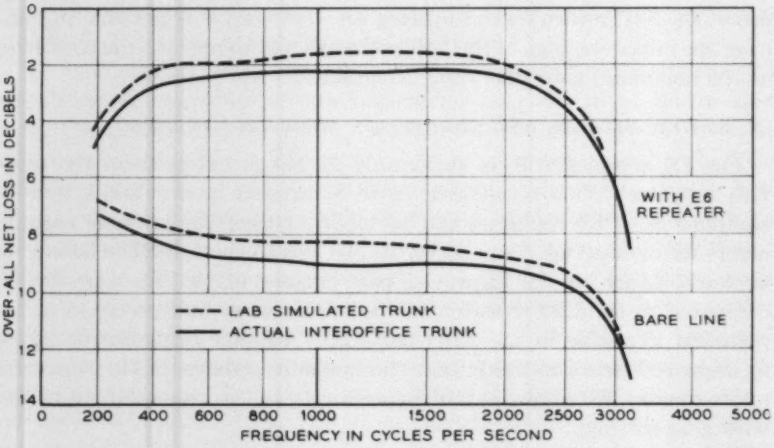
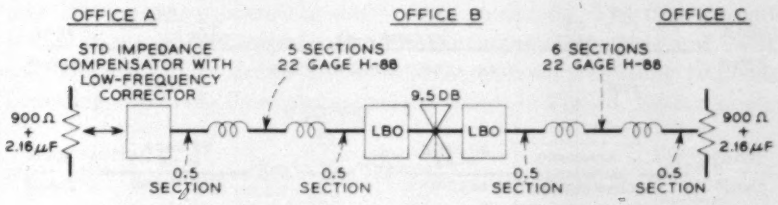


Fig. 43 — Performance of E6 repeater terminal on nine sections of 22-gage H-88 cable, frame to frame.

sufficient amount so as to remove most of the flat-with-frequency office losses at each end of the trunk.

A number of illustrations are given to show the performance of the E6 repeater. Fig. 43 is a comparison of test results obtained on a terminal repeater using physical and artificial cable. The repeaters were adjusted in such fashion as to maintain stability under all combinations of passive terminations at the trunk terminals. Where the structurals of the cable



MEASURED NET LOSS OFFICE A TO OFFICE C

1000 CYCLE		EFFECTIVE	
WITHOUT REPEATER	WITH REPEATER	WITHOUT REPEATER	WITH REPEATER
9.8 DB	1.5 DB	10.1 DB	2.5 DB

RETURN LOSS AT OFFICE A WITH COMPENSATOR NETWORK TERMINATION AT OFFICE C

500–2500 ∿	2000–3000 ∿	MINIMUM
19.5 DB	19.5 DB	14.5 DB AT 2500 ∿
WITH PRECISION NETWORK AT OFFICE C		
23.0 DB	23.0 DB	——

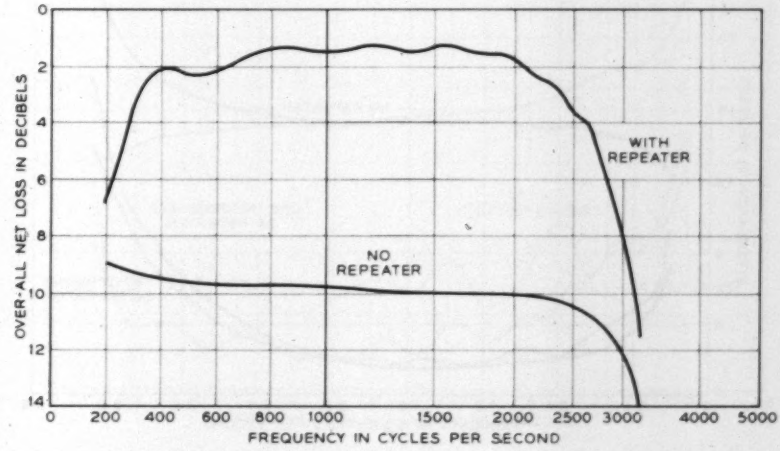
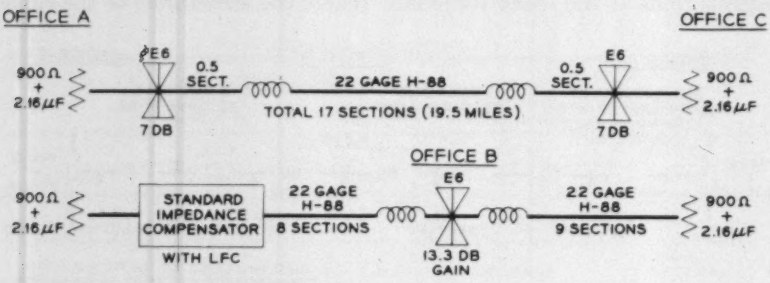


Fig. 44 — Performance of E6 repeater intermediate on 22-gage H-88 cable with impedance compensator.

were high, the results for both conditions are reasonably close and quite satisfactory.

On longer trunks requiring more than 6.5 db gain, an intermediate repeater with two LBO networks is required. Results on simulated cable are shown on Fig. 44. In this case a standard impedance compensator is used at the toll or office A end to improve the return loss.



RETURN LOSS AT OFFICE A

SWEEP	0.5 TO 2.5 KC	2-3 KC
2 REPEATERS	21.2 DB	21.8 DB
1 REPEATER	19.5 DB	20.5 DB

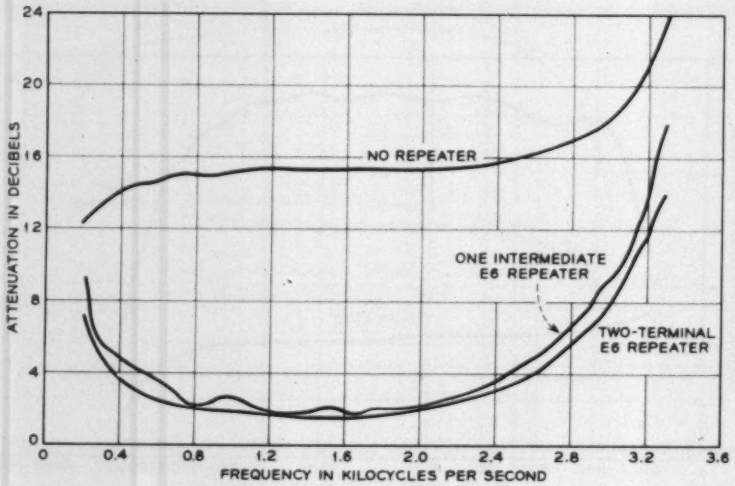


Fig. 45 — Over-all response with E6 repeaters at each end compared to a single intermediate E6 repeater.

For longer trunks with losses greater than 15 db, the best results are obtained with a terminal repeater at each end of the circuit. An intermediate repeater, giving somewhat inferior results to the two-repeater installation, may also be used if economies are desired. Results obtained for intermediate and two-terminal repeater operation, using simulated cable, are compared in Fig. 45. Return losses in either case are satisfactory when high cable structurals are available.

In the following examples, direct comparisons are made between E6 and E23 repeaters operating under field conditions. The first example is that of a tandem-completing trunk utilizing one repeater at an intermediate point. The circuit layout and the over-all frequency response, including effects of office equipment, are shown in Fig. 46. Each repeater

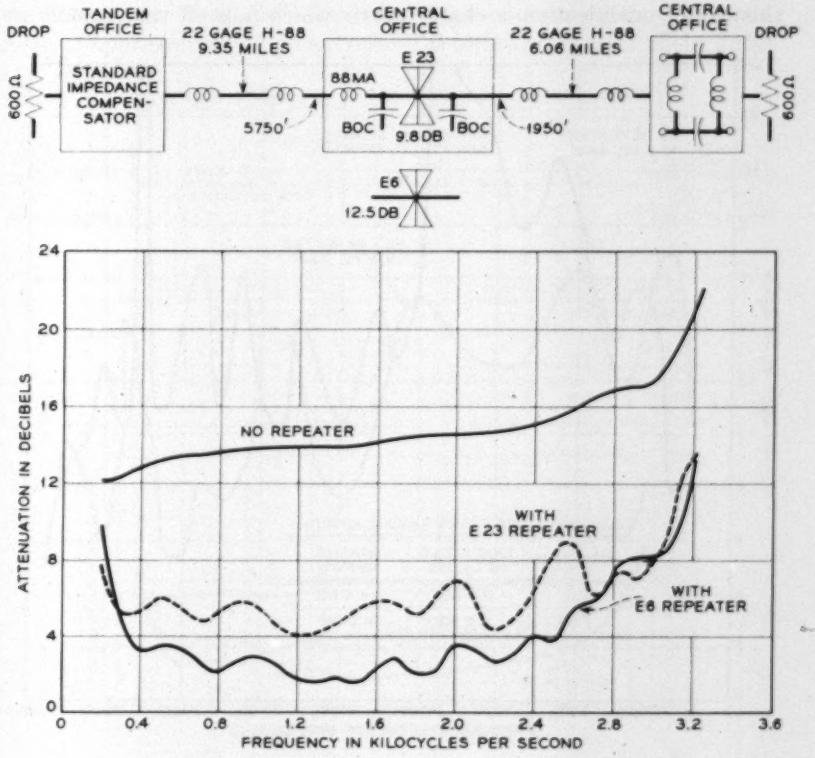
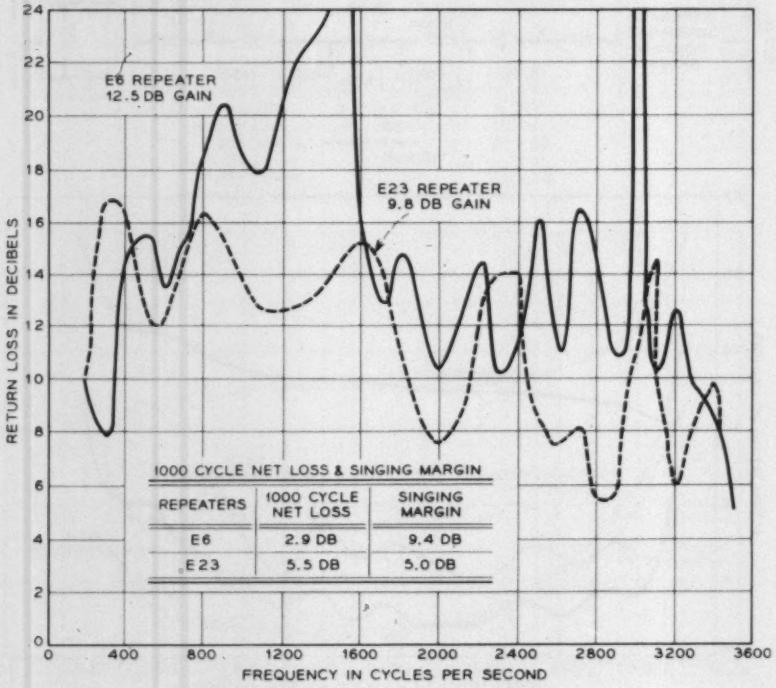


Fig. 46 — Over-all response including office equipment of tandem-completing trunk with intermediate E6 repeater.

was adjusted for maximum stable gain, using both idle and talking circuit terminations on the drop side of each office. Most of the office losses, amounting to approximately 1.3 db, were therefore removed. Stability of the repeater was also checked during the switching period to insure that singing did not occur during the interval required to establish a connection. The improvement of the E6 over the E23 is evident from the characteristics of Fig. 46. Measurements were made using 600-ohm apparatus and, as a result, certain irregularities in the response are to be observed. Had 900-ohm apparatus been used, lower net losses and smoother curves would have resulted due to the elimination of reflections. The E23 response is only as good as shown because a loading coil and a building-out capacitance equivalent to 0.5 end-section were added to the line on the 5,750-foot end-section side of the repeater. Without this circuit modification, a characteristic which falls off very rapidly at



1000 CYCLE NET LOSS & SINGING MARGIN

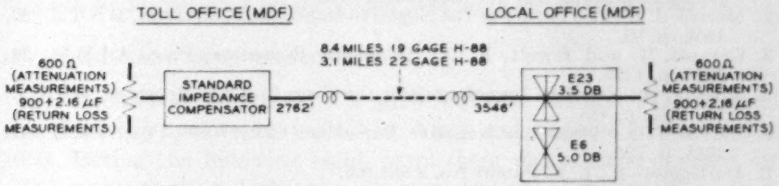
REPEATERS	1000 CYCLE NET LOSS	SINGING MARGIN
E6	2.9 DB	9.4 DB
E23	5.5 DB	5.0 DB

Fig. 47 — Return-loss characteristics of E6 and E23 repeaters as measured in field test.

the high end would have resulted. On the other side of the repeater some building-out capacitance was also required to bring the end-section up to 0.5-section. When the E6 repeater was used, however, no loading coil or external building-out capacitance was necessary, and each LBO was adjusted to full end-section with the elements available in the network.

In addition to the lower net losses made possible with the E6 repeater, better return losses and more adequate singing margins are also obtained. A plot of the return-loss characteristic taken under the field test conditions described above is shown in Fig. 47. The improved return loss response with the E6 repeater is apparent.

An example of a toll-connecting trunk terminal application is shown in Fig. 48. The repeater is located at the local or class 5 office. In this case, the trunk is made up of a mixture of 19- and 22-gage cable. Again, the E6 repeater gives a lower circuit net loss, amounting to an improvement of approximately 2 db over the E23 repeater. In spite of the differ-


TOLL OFFICE (MDF)
LOCAL OFFICE (MDF)
600 Ω (ATTENUATION MEASUREMENTS) 900+2.16 μF (RETURN LOSS MEASUREMENTS)
STANDARD IMPEDANCE COMPENSATOR
8.4 MILES 19 GAGE H-88
3.1 MILES 22 GAGE H-88
2762'
3546'
E23 3.5 DB
E6 5.0 DB
600 Ω (ATTENUATION MEASUREMENTS) 900+2.16 μF (RETURN LOSS MEASUREMENTS)

RETURN LOSS (TOLL OFFICE)

REPEATER	FREQUENCY RANGE		
	0.5–2.5 KC	2–3 KC	S.P. (DROP-TO-DROP)
E6	18.2 DB	15.7 DB	12DB
E 23	18.5 DB	15.7 DB	11DB

RETURN LOSS (LOCAL OFFICE)

REPEATER	FREQUENCY RANGE	
	0.5–2.5 KC	2–3 KC
E6	19.0 DB	16.0 DB

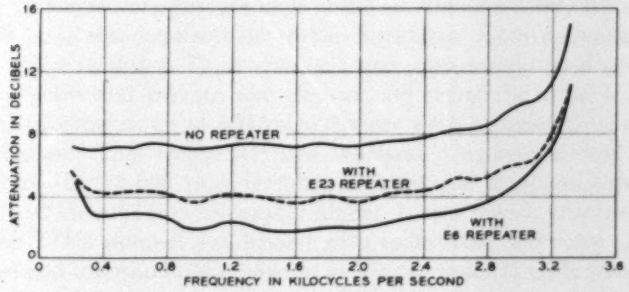


Fig. 48 — Over-all response of a toll-connecting trunk with terminal repeater at a local office.

ence in net loss, both repeaters gave about the same return loss results at the toll office end. At the repeater end, however, the E6 repeater gave considerably better return losses.

X. ACKNOWLEDGMENTS

The authors wish to acknowledge the many contributions made by their associates at Bell Telephone Laboratories, Western Electric Company and the American Telephone and Telegraph Company in the development of the E6 repeater. In particular, they wish to acknowledge the contributions of R. W. DeMonte to LBO network design techniques, W. C. Schmidt to packaging concepts and packaging development and J. K. Jones to equipment design and layout.

REFERENCES

1. Merrill, J. L., Jr., Rose, A. F. and Smethurst, J. O., Negative Impedance Telephone Repeaters, B.S.T.J., **33**, 1954, p. 1055.
2. Merrill, J. L., Jr., Theory of the Negative Impedance Converter, B.S.T.J., **30**, 1951, p. 88.
3. Gerardi, B. and Jewett, F. B., Telephone Repeaters, Proc. A.I.E.E., **38**, 1919, p. 1255.
4. Crisson, G., Negative Impedance and the Twin 21-Type Repeater, B.S.T.J., **10**, 1931, p. 485.
5. Linvill, J. G., Transistor Negative Impedance Converters, Proc. I.R.E., **41**, 1953, p. 725.
6. Darlington, S., U. S. Patent No. 2,663,806.
7. Fano, R. M., Theoretical Limitations on the Broadband Matching of an Arbitrary Impedance, J. Frank. Inst., **249**, 1950, pp. 57; 139.
8. Matthaei, G. L., Synthesis of Tchebycheff Impedance Matching Networks, Filters and Interstages, I.R.E. Trans., **CT-3**, 1956, p. 163.
9. Kuh, E. S., unpublished memorandum.
10. DeMonte, R. W., unpublished memorandum.
11. Guillemin, E. A., *Communication Networks*, Vol. 2, John Wiley & Sons, New York, 1935, p. 356.

Interstitial Channels for Doubling TD-2 Radio System Capacity

By H. E. CURTIS, T. R. D. COLLINS and B. C. JAMISON

(Manuscript received May 13, 1960)

The TD-2 microwave radio relay system at the present time provides six broadband channels in each direction of transmission in the frequency band between 3700 and 4200 mc, each channel being capable of transmitting several hundred voice circuits. This paper describes arrangements and some of the technical problems involved whereby the number of broadband channels can be doubled with no significant penalty to the existing channels.

I. USE AND GROWTH OF TD-2 SYSTEMS

TD-2 routes were installed at a rapid rate after the first TD-2 system[1] was placed in service between New York and Chicago in September 1950. During the following eight years there was a large demand for additional intercity television, message and private line circuits to meet the expanding business requirements. Because the TD-2 system gives excellent transmission performance and also has economic advantages over other systems, especially where new routes are needed, a large part of the additional Bell System intercity facilities provided during the eight-year period consisted of TD-2 channels.

Long video network circuits were obtained almost entirely from TD-2 channels, and side connections to the backbone video networks were usually TD-2 facilities. These channels were also widely used to obtain intercity voice and private line circuits. By using the type L carrier terminal equipment, up to 600 voice circuits were assigned to one TD-2 radio channel. This technique uses baseband frequencies only up to about 3 mc for the 600 voice circuits, although the baseband capability of the TD-2 channel is considerably higher. It was found, however, that, when the TD-2 channel was loaded with more than 600 voice circuits, the noise and intermodulation distortion on the voice circuits were often too high to meet circuit objectives.

By the end of 1958 a microwave network had been established which connected together most of the large cities in the United States and Canada, as shown in Fig. 1. This network was comprised of over 30,000

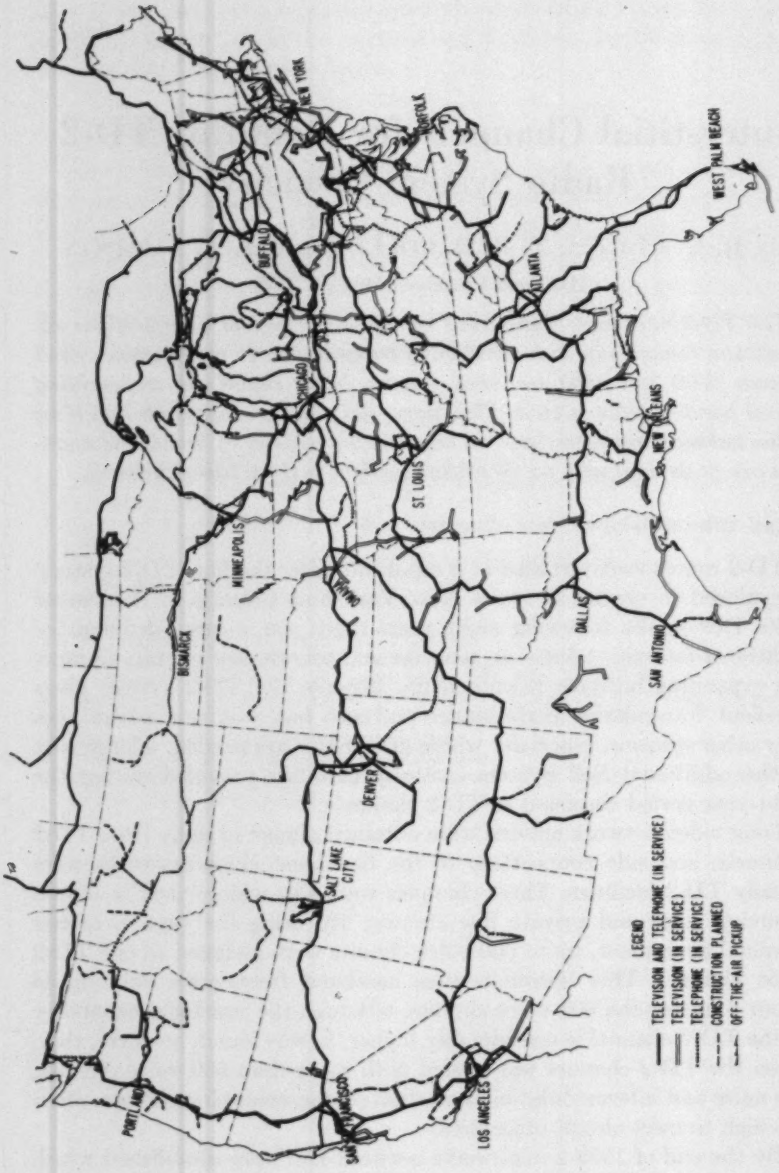


Fig. 1— TD-2 routes in 1958.

TD-2 route miles between some 1300 individual stations scattered throughout the country. These routes provided about 67,000 miles of intercity television circuits and about 20,000,000 telephone circuit miles. The TD-2 radio channels installed over the approximately eight-year period comprised about 80 per cent of all intercity video and about 35 per cent of all long distance telephone circuit mileage in the Bell System plant.

II. NEED FOR ADDITIONAL TD-2 CHANNELS

Before 1958 many of the existing backbone routes had all six channels in at least one direction of transmission either in service or planned for service; it was becoming increasingly apparent that six radio channels would not be adequate to take care of the wideband channel needs, and that construction of parallel backbone routes would be necessary.

Television requirements for circuits resulted in some routes having more channels in service in one direction of transmission than in the other direction. These unbalanced requirements, together with other needs for two-way circuits, resulted in a number of routes reaching their capacities in at least one direction of transmission.

Additional TD-2 channels could be obtained in some circumstances by constructing new TD-2 routes, and this was done where practicable. Such new routes required new radio towers and buildings to house the equipment and power plants. This was expensive, particularly where the geographical length of the routes had to be increased in order to avoid interference to existing routes.

Other radio systems which would use frequencies in other common carrier bands and give high-quality circuits meeting long distance transmission requirements were not yet available. The new TH microwave system is expected to be available for service about the end of 1960, and each individual TH broadband channel will provide about three times the circuit capacity of a TD-2 channel.[2] The TH channels will usually be installed along routes where their high telephone circuit capacity is needed and where higher costs per radio channel will be justified.

The possibility of adding interstitial channels between the regular TD-2 channels had been considered, in view of the unused 20-mc interstitial frequency space available between channels, but a study revealed that it was not practicable with the delay lens antenna and associated rectangular waveguide system.[3] Starting in 1955, the horn-reflector antenna with the associated circular waveguide system became standard for all new installations, and it then became possible to cross-polarize signals between adjacent TD-2 channels to obtain additional discrimina-

tion between adjacent channel signals.[4] By the beginning of 1958 the necessity of providing more circuits along fully loaded TD-2 routes resulted in the investigation of the feasibility of adding interstitial channels on TD-2 routes.

III. DESCRIPTION OF TD-2 SYSTEM

The TD-2 system has been described extensively in the literature, but, for the purpose of the present paper, certain features will be reviewed briefly.[5] Each baseband channel has a frequency capacity extending from about 30 cycles to 6 to 10 megacycles per second depending on the length of the system. Frequency modulation is employed to raise each channel to an intermediate frequency of 70 mc, from which each is shifted to its proper location in the microwave frequency spectrum. The frequency plan for a typical repeater point is shown in Fig. 2.

Channels are so assigned that there is always a spacing of 40 mc between the carrier of any transmitter operating in a given direction and

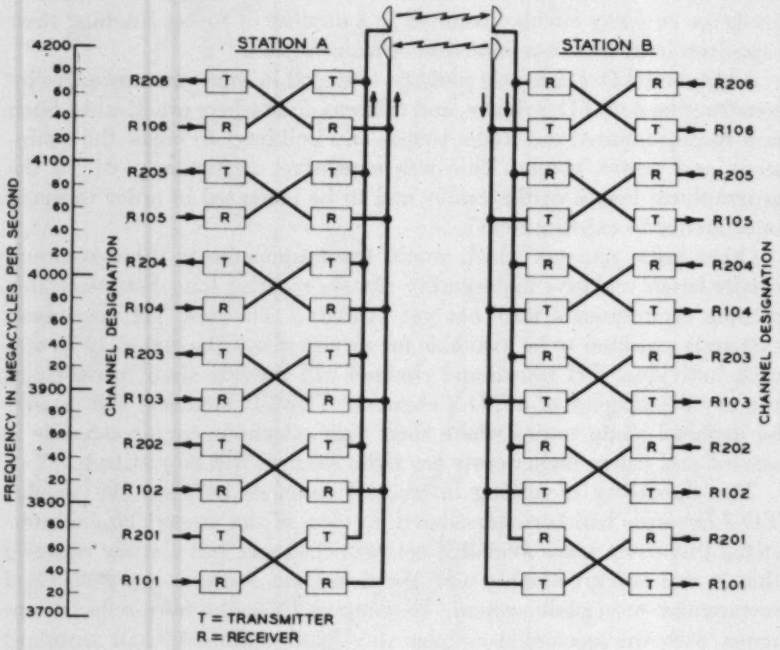


Fig. 2 — Frequency plan for six channels.

the carrier of the nearest receiver operating in the same direction. Since each FM channel potentially occupies a total band of 20 mc, there is a total of six interstitial bands of 20 mc each in each direction in which no significant quantity of energy is present.

Successful operation of the TD-2 system requires freedom from interference between microwave channels under all reasonable conditions. The potential interference which does exist arises from radio-frequency coupling paths. These couplings may, for convenience, be separated into two types. The first, which is called near-end, is due to high-level energy radiated from one antenna on a tower falling into receivers connected to the receiving antenna on the same tower. In the frequency plan shown in Fig. 2, the significant interferences of this type are 40 mc removed from the center of the band of the disturbed channel. The effect of this RF interference is minimized by

(a) the transmission loss between the two antennas,

(b) the selectivity of the RF channel-dropping filters and image-suppression filters,

(c) the discrimination of the IF amplifiers.

The net loss, including antenna gains between horn-reflector antennas side by side and pointed in the same direction, is in excess of 95 db in the absence of foreground reflections. However, measurements at a large number of locations show that the effect of foreground reflections is to distribute the loss over a wide range of values. Fig. 3 shows a cumulative probability distribution curve of 97 measurements of side-to-side coupling loss between horn-reflector antennas. Even the poorest value shown, 61 db, is adequate for six-channel operation. The RF filters offer about 16 db of loss at frequencies 40 mc removed from the center of the channel band. The main IF amplifier, however, provides about 60 db of discrimination at 30 mc and 110 mc relative to 70 mc.

Since the normal transmitting power as measured at the input to the transmitting waveguide is approximately 0.5 watt, or +27 dbm, and the received carrier power at the converter during nonfading conditions is −35 dbm for the average path, the level difference between these two points on the same tower is 62 db. However, the sum of the above enumerated losses, together with those in the IF preamplifier, the transmitter modulator and the first stage of the transmitting RF amplifier, insures complete freedom from near-end interference between microwave channels under all normal conditions.

The second type of RF interference, called far-end, is the result of energy intended for one receiver at a particular point falling into another receiver at the same point. In this case the nearest disturbing channel

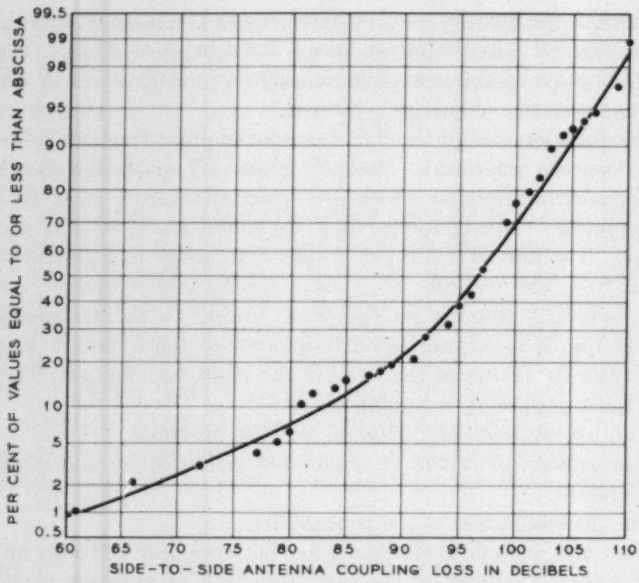


Fig. 3 — Side-to-side antenna coupling loss.

is 80 mc removed from the center of any particular disturbed channel. For example, at Station B in Fig. 2 the carrier from Station A at 4090 mc may, potentially at least, interfere with the similarly directed channels at 4170 mc and 4010 mc. Radio-frequency filters are used which provide about 50 db of loss to frequencies 80 mc removed from the channel operating frequency; here again the sum of all the circuit losses at side frequencies is sufficient to insure freedom from far-end interference.

IV. INTERSTITIAL CHANNELS

Six additional two-way microwave channels may be derived by utilizing the interstitial bands mentioned above. A frequency plan for such operation is shown in Fig. 4, in which the six additional channels are shown as dotted lines. It will be noted that here adjacent receiving channels are only 20 mc apart and, furthermore, the minimum difference in frequency between transmitters and receivers on the same tower is also 20 mc.

Inasmuch as the total discrimination of the repeater to frequencies 20 mc from midband is not great, supplementary means had to be found

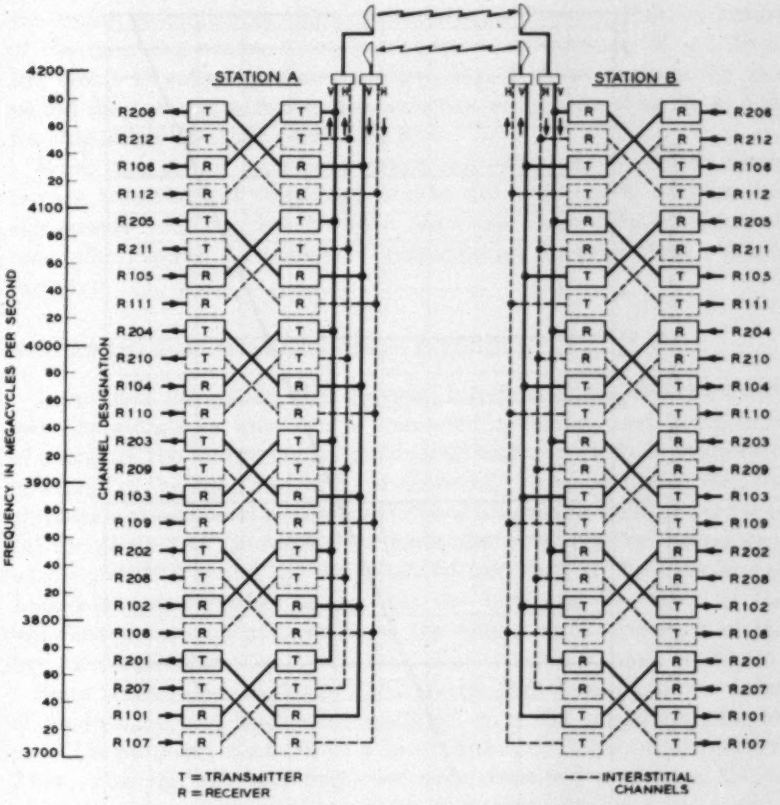


Fig. 4 — Frequency plan for 12 channels.

for improving the discrimination between such closely spaced channels. For far-end interference between similarly directed channels only 20 mc apart, additional discrimination may be obtained by transmitting one set of six channels with vertical polarization and the other set of six channels with horizontal polarization, as shown symbolically in Fig. 4. The two sets of oppositely polarized channels are separated at the receiving waveguide by means of a suitable network.

It is essential that during fading the amplitude of the vertical component of any particular horizontally polarized carrier shall not become excessive relative to the amplitude of the vertically polarized carrier located 20 mc away, and vice versa. Cross-polarization during fading

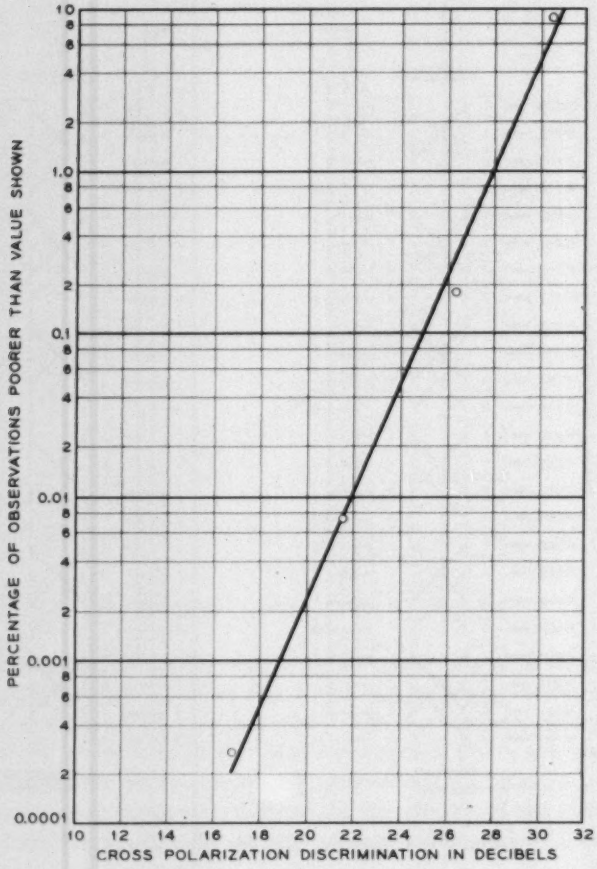


Fig. 5 — Cross-polarization discrimination.

periods was studied experimentally* on a 23-mile path between Murray Hill, New Jersey, and Holmdel, New Jersey. This was done by transmitting a vertically polarized carrier at 4008 mc and a horizontally polarized carrier at 3980 mc, and observing them with a receiver tuned to each but with both receivers arranged to accept only vertically polarized waves. Fig. 5 is a cumulative distribution curve of the instantaneous difference in level between the two received components obtained in

* This experimental work was carried out by G. M. Snow, formerly of Bell Telephone Laboratories.

the month of September 1954. It should be pointed out that, by nature of the recording means, there is an inherent uncertainty of ±5 db in the observed values of discrimination. Fig. 5 shows that during this period the cross-polarization discrimination was poorer than 20 ± 5 db for only about 0.002 per cent of the time.

While the use of cross-polarization materially improves the interference from far-end couplings, it does not appreciably increase the side-to-side coupling loss between antennas; consequently, near-end crosstalk becomes an important consideration in interstitial channel operation.

V. INTERCHANNEL INTERFERENCE — THEORY

In an ideal FM system, the baseband interference due to a relatively weak interfering FM wave can be computed by beating each component of energy of the interfering (or disturbing) carrier with each component of energy of the interfered-with (or disturbed) FM wave, noting that the amplitude at baseband of any interference component is proportional to (a) the product of the respective amplitudes of the two RF components which produced it and (b) the baseband frequency of the interference component; and further noting that the baseband frequency of the interference component is equal to the difference in frequency of the two RF components.

Since transmission over the TD-2 system uses a relatively low index of modulation, the sidebands associated with the carrier are, for all practical purposes, confined to a band within ±10 mc of the carrier. Thus, with the standard frequency plan described in Section III, in which the nearest disturbing carrier is 40 mc from the disturbed carrier, any interference between the two would fall above a baseband frequency of 20 mc.

If channels were operated in the interstices available under the present frequency plan, potential interference would extend from zero baseband frequency upwards, having maximum power at 20 mc. A quantitative analysis using this theory shows that, for the types of baseband signals normally transmitted over the TD-2 system, the interference falling in the useful baseband frequency range of the TD-2 system would be expected to be negligible if the disturbing carrier at the FM detector were always at least 6 to 10 db below the disturbed carrier.

Preliminary laboratory studies and published information indicated that the interference at baseband was far higher than expected and, furthermore, did not obey the laws indicated by ideal simple FM theory.[6]

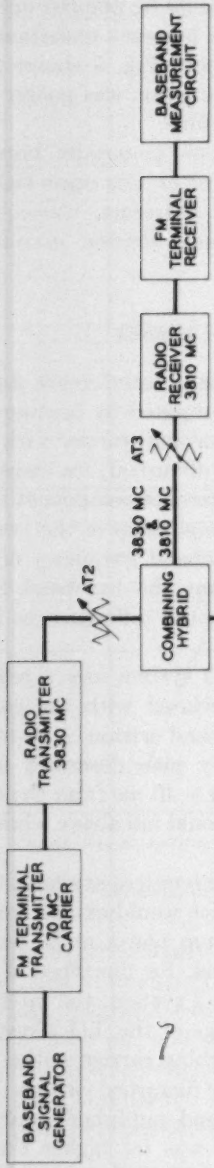


Fig. 6 — Laboratory test arrangement.

For this reason a detailed investigation to establish feasibility of interstitial operation was implemented in the laboratory and continued in the field.

VI. LABORATORY STUDY

Some of the characteristics of direct adjacent channel interference between TD-2 channels 20 mc apart had been studied briefly in the laboratory as early as 1952. The desire to operate the TD-2 system as a 12-channel system using this spacing made it necessary to extend the previous work and to produce an engineering plan for successful operation of the 12-channel system.

A typical laboratory circuit arrangement for testing such a proposal is shown in Fig. 6 in block schematic form. The upper left-hand part of the figure shows the disturbing channel, operating at 3830 mc and modulated by a signal from a baseband signal generator. The equipment in the lower left section provides the 3810 mc carrier that represents the disturbed channel. In the early part of the studies this carrier was unmodulated, as indicated in the figure. The two microwave carriers are combined in the hybrid junction and delivered to the radio receiver, which is tuned to the carrier frequency of the disturbed channel. The 70 mc output of the radio receiver is passed to the FM receiver, which is in turn followed by baseband measuring equipment of appropriate type. The variable attenuator AT1 is used to adjust the 3810 mc carrier power at the input to the radio receiver to any desired value. Similarly, AT2 is used to adjust the received power of the 3830 mc disturbing carrier. The difference between the decibel settings of the two attenuators represents the cross-polarization discrimination, and this proved to be the significant parameter in many of the tests.

Tests were made using single-frequency baseband tones to modulate the disturbing channel while interference in the disturbed channel was observed at the output of the FM receiver connected to the 3810 mc radio receiver. Results of these tests are summarized broadly as follows:

1. Interference appeared in the disturbed channel at exactly the same baseband frequency as was applied to the other channel.

2. A change of 1 db in the carrier ratio of the two channels produced a change in the baseband interference in the disturbed channel of 2 db, provided the disturbing carrier was weaker than the disturbed carrier.

3. A change of 1 db in the level of the baseband input to the disturbing channel produced in general a change of 1 db in the interference observed in the disturbed channel output.

4. The baseband interference was observed to be essentially independ-

ent of the modulation frequency on the disturbing carrier when this frequency was in the 50 kc to 5 mc range.

5. With a fixed difference in level between the disturbing and disturbed carriers, the interference as observed at baseband was independent of the power of the disturbing carrier at the input of the radio receiver.

Certain exceptions to these generalizations were noted but are not considered to be essential to the argument. For example, with low-baseband modulating frequencies, e.g., 50 kc, a change in the frequency deviation from ± 4 mc to ± 0.4 mc produced a change in interference in the disturbed channel of 26 db rather than 20 db.

Fig. 7 shows some of the characteristic relations between the FM carrier ratios and the baseband interference. The abscissa is the ratio in decibels of the two carrier powers as measured in the receiving waveguide. The ordinate is the observed baseband interference in decibels on an equal-level crosstalk scale.* Over a considerable range of carrier ratios the slope of the curves is essentially 2:1. However, as the carrier ratio begins to approach unity, the slope is considerably greater than 2:1. Since a working system will not be satisfactory when the interference is large, the region where the disturbing carrier is stronger than the disturbed carrier is only of academic interest. While the shapes of the curves are almost the same over the region of interest, the position of the curves with respect to the carrier ratio depends upon the discrimination against the disturbing channel that is provided by the frequency characteristic of the receiver in the disturbed channel. The experimental work indicated that the crosstalk mechanism which is characteristic of adjacent channel interference exists in the limiter of the FM receiver and also at any point in the system where there is a pronounced tendency to compress the signal. Such a mechanism has been suggested in an unpublished memorandum by W. R. Bennett of Bell Telephone Laboratories. This mathematical work shows that, when an unmodulated FM carrier and a weaker sinusoidally modulated FM wave are passed through an ideal clipper limiter, distortion is introduced on the stronger FM carrier, some of which appears at baseband frequencies as interference having the same frequency as that of the modulation on the weaker carrier. The amplitude of this interference varies as (a) the square of the ratio of the amplitude of the two carriers and (b) the square of the ratio of the voltage at which the limiter clips to the signal voltage impressed upon it.

* On this scale the number -40 means that the measured interference at baseband is 40 db below the value of the baseband tone which would be measured at the output of an FM receiver on the disturbing channel.

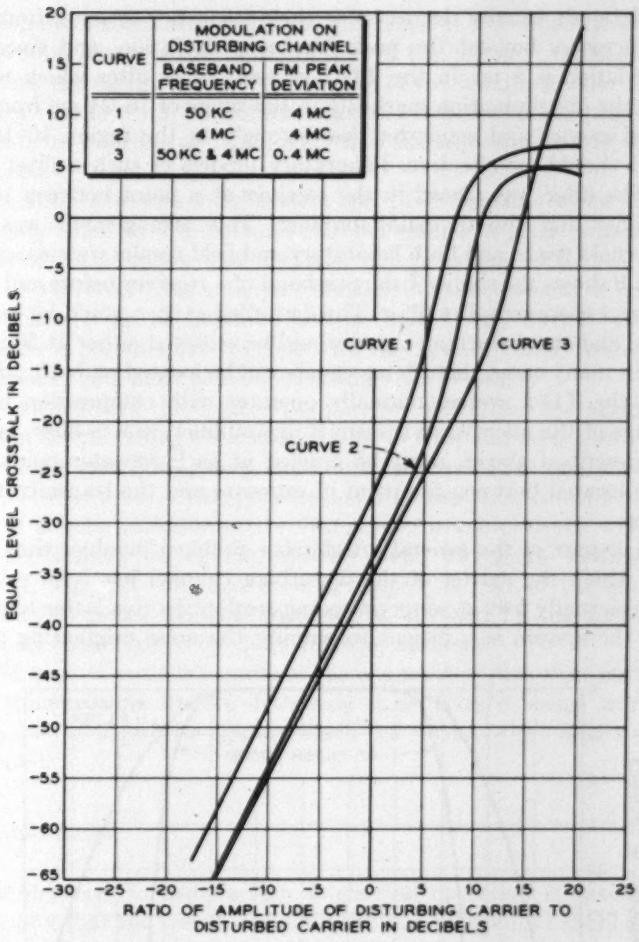


Fig. 7 — Crosstalk between microwave channels — laboratory measurements.

The next series of experiments was undertaken to determine how the frequency discrimination at the edges of the channel passband might be changed to hold the interchannel interference to tolerable limits in the presence of a limited amount of cross-polarization discrimination. It was found that sufficient improvement could be obtained if the frequency discrimination of the disturbed channel was increased in the region of the disturbing carrier and its signal sidebands. Since the strength of

these sidebands usually decreases as their frequency spacing from the carrier increases beyond the peak frequency deviation, and since the peak deviation is 4 mc in the TD-2 system, an IF filter which would increase the discrimination markedly in the range of 16–24 mc from the disturbed carrier and somewhat less strongly in the region 10–16 mc removed should be effective. Laboratory models of such a filter were made. The filter was placed in the receiver at a point between the IF preamplifier and the IF main amplifier. This arrangement was also tested in field trials, and both laboratory and field results were encouraging. Fig. 8 shows the shape of the passband of a receiver before and after adding an experimental IF filter. The disturbed carrier would be located at 70 mc and the disturbing signal would be centered either at 50 mc or 90 mc. In many cases, disturbing signals will be located on both sides.

Since the TD-2 system normally operates with compression in the final stage of the microwave transmitting amplifier, this IF filter, for the reason described above, must be applied at each repeater point, and must be located between the point of exposure and the transmitting RF amplifier.

Another part of the general interference problem involves the condition in which the carrier of the disturbing channel has been severely faded or actually lost at some preceding section. In this latter case, the gain of the system is sufficient to amplify the noise originating in the

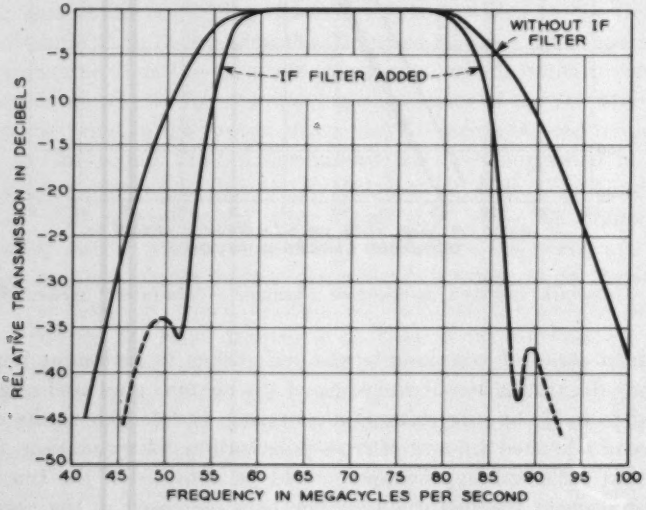


Fig. 8 — Radio receiver gain frequency characteristics.

converters until the transmitter output consists of a broad band of noise whose total power is substantially that of the carrier it replaces. In this situation the energy density at the edge of the band is greater than it is when a carrier with its normal sideband distribution is present. The tendency to interfere or "spill over" into the adjacent channel is greatly increased.

The automatic switching system for TD-2 operates when the high-frequency noise in a broadband channel reaches a predetermined level as a result of a fade or loss of carrier in a preceding section of the system.[7] If spill-over from a failed interstitial channel occurs, a nominally good channel can be impaired and an unnecessary switch will ensue. The filter is not adequate to cope with this situation and, consequently, the repeater gain must be reduced during such periods.

The elements of the engineering plan for interstitial operation can be found in the preceding discussion. A new IF filter and means for reducing the repeater gain during deep fades must be provided at each repeater on an interstitial route. Cross-polarization discrimination is available in the form of a polarization-sensitive 4-kmc microwave network which connects rectangular waveguide into the round waveguide leading to the antenna. This network had been developed as part of the equipment necessary to permit systems operating at 4, 6 and 11 kmc to be simultaneously connected to one horn-reflector antenna.[8]

Field tests of various aspects of the engineering plan were carried out from time to time. Before discussing these tests it seems worthwhile to discuss the objectives against which the results of the field tests were evaluated.

VII. OBJECTIVES

The objective for satisfactory interstitial operation is based on the philosophy that the speech crosstalk from an adjacent channel during a fade should be well masked by the first circuit noise simultaneously present. For systems engineering purposes, this objective must be interpreted in terms of a required ratio of disturbed carrier to disturbing carrier.

In the case of near-end crosstalk as defined above, the desired received carrier is subject to normal fading whereas the disturbing carrier is not. For this condition it has been determined experimentally that, without the special IF filter, if the disturbing carrier is 35 db or more below the disturbed carrier during conditions of free-space transmission, then, during fading, the baseband interference from the interstitial channel

into the disturbed channel will always be below the noise in the disturbed channel.

In the case of far-end crosstalk, fading of the disturbing carrier is correlated to a degree with fading of the disturbed carrier, and this in turn leads to an objective somewhat less than 35 db. The limited amount of data at hand indicates that for this case a ratio of about 20 db is adequate. If the fading on the two channels is uncorrelated, the objective will be 35 db for far-end crosstalk as well as for near-end crosstalk.

Fig. 9, taken from Ref. 9, shows how the fluctuation noise and speech

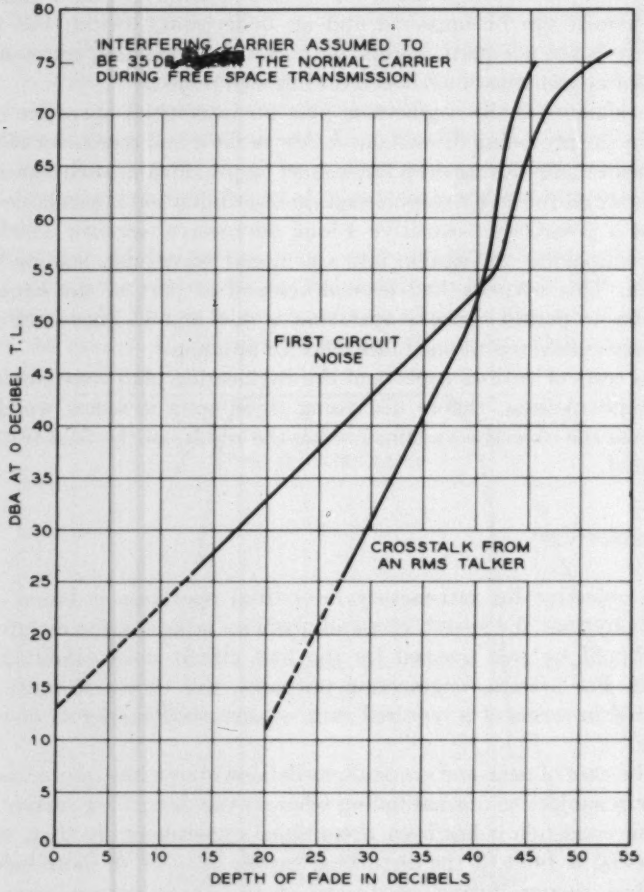


Fig. 9 — Expected noise and crosstalk vs. fade.

crosstalk contributions of one repeater may be expected to increase with depth of fade. Talker volumes are distributed over a wide range, and the volume assumed here is sometimes referred to as the volume of the "rms" talker. This is approximately 10 db below the loudest talker. At a 35 db fade, the point at which service is normally transferred to a nonfaded channel, the speech crosstalk is 7 db below the noise power.

It is also convenient for some purposes to specify the above objective in terms of equal-level crosstalk at baseband frequencies. The conversion can readily be made by means of the experimental relationship shown in Fig. 7. Thus, if the 35 db objective were met, then under free-space conditions the equal-level crosstalk would be completely negligible. For a 35 db fade, the carrier ratio would be 0 db and the corresponding equal-level crosstalk objective would be −35 db.

VIII. FIELD TESTS

In the six repeater sections between West Unity, Ohio, and Grant Park, Illinois, interstitial channels had already been installed, thus providing a suitable location for feasibility field tests. Channels R201 and R207 in the east to west direction and R101 and R107 in the opposite direction (see Fig. 4 for channel nomenclature and frequencies) were used for the tests since R101 could interfere with R107 and produce far-end crosstalk, and similarly, for example, it could couple into R207 and produce near-end crosstalk. Both crosstalk paths involve a spacing of 20 mc.

The following basic tests were among those made:

1. *Sine-Wave Crosstalk*

A sine wave of suitable power was applied to the baseband input of R201, for example, and the magnitude of the crosstalk at this same frequency was observed at West Unity at the baseband outputs of R107 and, at Grant Park, of R207. This was carried out for various values of simulated radio frequency fades at each repeater section on R107 and R207.

2. *Video Crosstalk*

The test detailed above was repeated using a video signal. In this case the outputs of R107 and R207 were examined on a video monitor for evidence of picture crosstalk for simulated fades, as described above.

3. *Noise Spill-Over*

The carrier was removed from one radio channel at the head end of the system, the gain of the subsequent repeaters thereby rising, as described previously, until the repeater output consisted of a broad band of noise. The magnitude of the spill-over of this noise into the adjacent channel was observed for various depths of fades.

4. *Loss of Fading Margin Due to Noise Spill-Over*

In a TD-2 system, one channel out of six is normally set aside as a protection channel. In each operating channel, noise is sensed at a baseband frequency of about 9 mc and, if the noise is excessive, a switch is made to the protection channel if it is available. Noise spill-over increases the noise at the upper video frequencies and hence causes a faded circuit to ask for the protection channel sooner than it otherwise would. This, in turn, ties up the protection channel longer than is actually necessary and, as a result, increases the amount of time a switch cannot be made because the protection channel is unavailable.

Fig. 10 shows typical results of sine-wave crosstalk measurements in which the modulating tone on the interfering channel is 3.59 mc and various repeater sections are faded. The coupling path in this case is between similarly directed channels and generates far-end crosstalk. The results are plotted in terms of equal-level crosstalk.

Using the more severe of the two objectives suggested above for far-end crosstalk, it was shown that this led to an objective of −35 db for equal-level crosstalk loss for a 35 db fade. It is apparent then that three out of the six far-end coupling paths failed to meet this requirement.

The addition of the supplementary IF filter substantially reduced the magnitude of the crosstalk, as shown in Fig. 11. In this case the poorest equal-level crosstalk value for a 35 db fade was −35 db, which meets the stated objective.

Observations showed video crosstalk mixed with the noise for deep fades, but this was eliminated by the application of the supplementary IF filter.

Figs. 12 and 13 show the effect of noise spill-over from channel R201 into R207 as the latter is artificially faded at Kouts. The first of these two figures shows the measured noise spectrum on R207 with R201 operating in normal fashion. These curves exhibit satisfactorily the expected 6 db per octave increase of noise with baseband frequency. Fig. 13, however, shows the effect of removing the carrier from R201

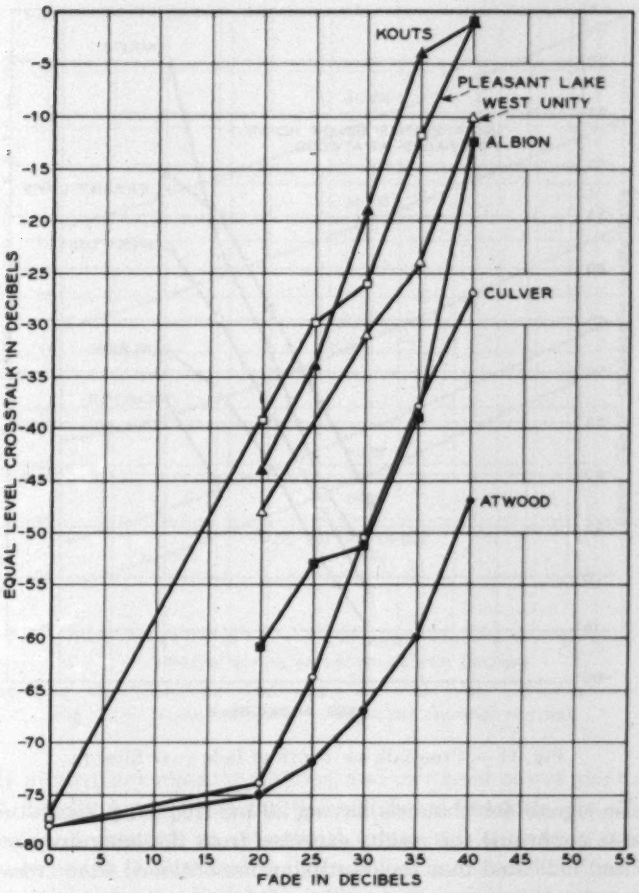


Fig. 10 — Crosstalk vs. depth of fade — IF filters removed

at West Unity. The effect of noise spill-over from R201 is most marked in the region above 6 mc.

IX. APPLICATION TO EXISTING ROUTES

The application of interstitial channels to an existing TD-2 route requires the use of the horn-reflector type of antenna and associated circular waveguide system in order to permit the cross-polarization of

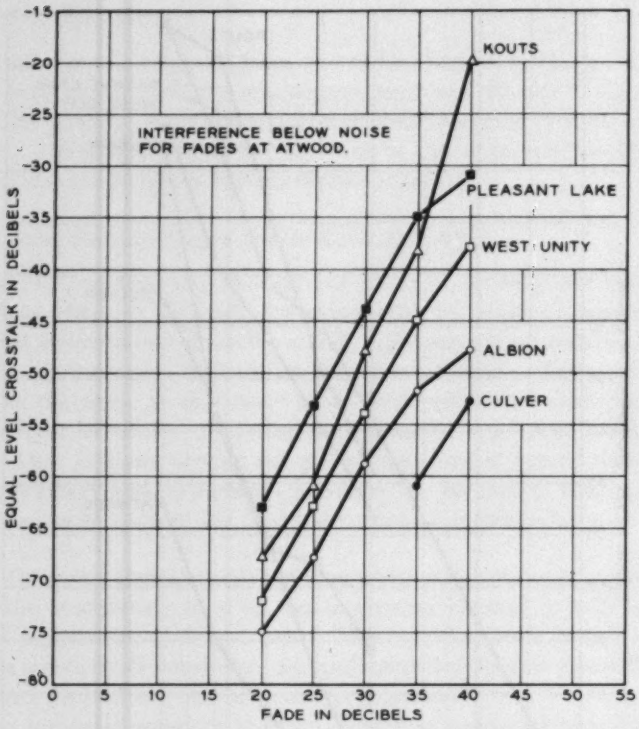


Fig. 11 — Crosstalk vs. depth of fade — IF filter in.

the radio signals for channels having 20 mc frequency separation. The field tests confirmed the results expected from the laboratory measurements and indicated that the discrimination obtained from cross-polarization and IF filters would provide satisfactory discrimination against adjacent channel signals. Existing delay-lens antennas with their associated rectangular waveguides must be replaced and coupling networks must be installed at the bottom of the tower for combining the cross-polarized signals. Temporary arrangements, such as parabolic antennas, may be necessary in order to keep in service the circuits assigned to the existing TD-2 channels and to avoid interruptions during the cutover.

It is planned to operate the interstitial channels added on a route as a system separate from the existing regular channels. The interstitial channels will have their own protection channels, because adjacent channels 20 mc apart will frequently fade at the same time and simultaneously need the same protection channel. Building additions, par-

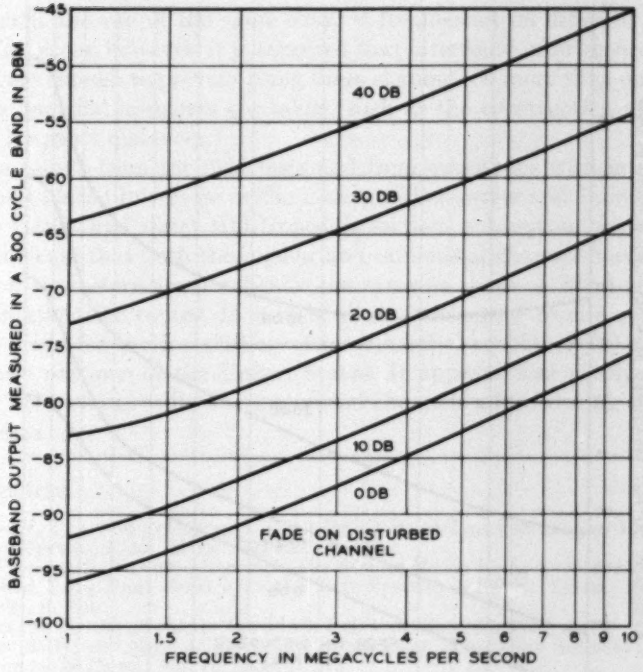


Fig. 12 — Noise spill-over — adjacent channel normal.

ticularly at auxiliary repeater stations, and increased power plant capacity will usually be required for the additional channels. In all other respects TD-2 channels on interstitial routes will operate in a normal manner, since the IF filter and the gain reducing means mentioned in an earlier section are essentially the only additional circuit arrangement required on an interstitial channel route.

Some of the factors affecting the decision as to application of interstitial channels to an existing TD-2 route are the circuit growth needed, the freedom from interferences arising from parallel or branching routes using the same interstitial channel frequencies and the type of antennas already installed. In some cases there may be interference from other radio routes such as sidelegs or converging routes which might limit the number of interstitial channels that could be added.[9]

TD-2 routes having horn-reflector antennas are particularly attractive candidates for interstitial channel applications from an economic standpoint. The cost of replacing the delay-lens antennas and rectangu-

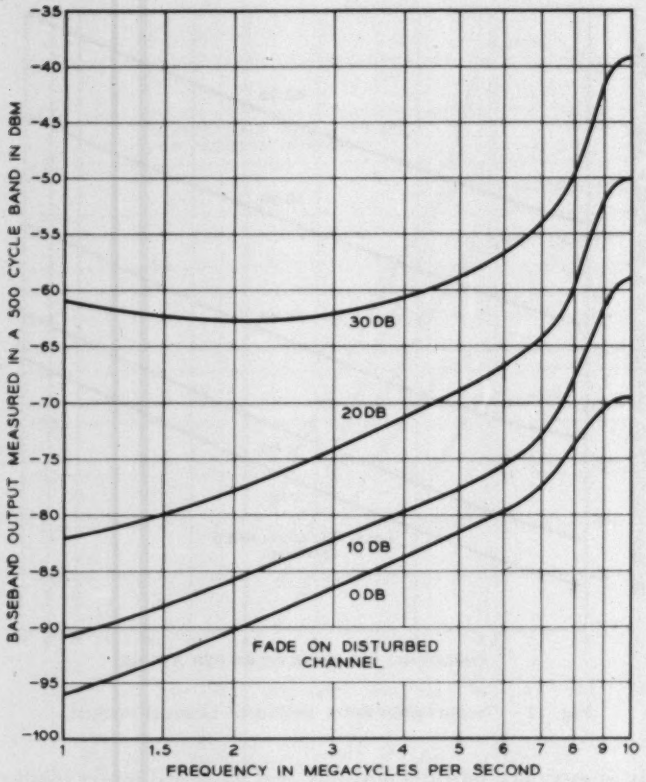


Fig. 13 — Noise spill-over — adjacent channel open.

lar waveguides by horn-reflector antennas and circular waveguides, and at the same time maintaining existing service, is substantial.

In general, the same factors apply for sideleg routes as for main through routes. However, in sideleg route cases it will be necessary to make sure that the addition of interstitial channels on the sidelegs will not restrict the number of such channels which could be added on main routes because of possible frequency interferences.

X. CONCLUSION

It appears that widespread use of TD-2 interstitial channels is possible, since the results of the field trial indicate that the interference between nearby and branching TD-2 routes would in general be sufficiently low

to permit the use of the same channel frequencies on different routes. In a few cases, however, it is expected that interroute interference might be severe enough to prevent using these channels on more than one route unless remedial measures are taken, such as the construction of a new route for short distances.

The results from the field tests and from experience with interstitial channels placed in service in the Chicago-Mishawaka and Indianapolis-Terre Haute and Troy Hill-Hopedale sections subsequent to the field trial indicate that both the regular and interstitial channels have transmission characteristics practically the same as those of TD-2 channels on noninterstitial routes. In view of these satisfactory results, plans are being made for the installation of interstitial channels on TD-2 routes in many portions of the United States. It appears that nearly all the main TD-2 routes will have interstitial channels added during the next few years.

REFERENCES

1. Grieser, T. J. and Peterson, A. C., A Broad-Band Transcontinental Radio Relay System, Elect. Engg., **70**, 1951, p. 810.
2. McDavitt, M. B., 6000-Megacycle-Per-Second Radio Relay System for Broad-Band Long-Haul Service in the Bell System, A.I.E.E. Trans., **76**, Pt. 1, 1958, p. 715.
3. Lince, A. H., Antennas for the TD-2. Bell Lab. Rec., **30**, 1952, p. 49.
4. Friis, R. W. and May, A. S., A New Broad-Band Microwave Antenna System, Comm. & Elect., No. 35, 1958, p. 97.
5. Roetken, A. A., Smith, K. D. and Friis, R. W., The TD-2 Microwave Radio Relay System, B.S.T.J., **30**, Pt. 2, 1951, p. 1041.
6. Medhurst, R. G., Hicks, Mrs. E. M. and Grossett, W., Distortion in Frequency Division Multiplex FM System Due to an Interfering Carrier, Proc. I.E.E., **105**, Pt. B, 1958, p. 282.
7. Welber, I., Evans, H. W. and Pullis, G. A., Protection of Service in the TD-2 Radio Relay System by Automatic Channel Switching, B.S.T.J., **34**, 1955, p. 473.
8. Harkless, E. T., A Network for Combining Radio Systems at 4, 6 and 11 KMC. B.S.T.J., **38**, 1959, p. 1253.
9. Curtis, H. E., Radio Frequency Interference Considerations in the TD-2 Radio Relay System, B.S.T.J., **39**, 1960, p. 369.

A Study of Talking Distance and Related Parameters in Hands-Free Telephony

By MARK B. GARDNER

(Manuscript received December 16, 1959)

This paper outlines the problems of providing satisfactory hands-free operation of the telephone and discusses various methods which can be applied to their solution. Particular attention is given to acoustic environment, proximity talking (as applied to hands-free operation of the telephone) and voice switching. Preference indications were obtained for 18 subjects in 18 different locations for proximity and distant talking without voice switching, and for the same number of subjects and locations for proximity and distant talking with voice switching. In the latter tests, supplementary data were also obtained without voice switching. Regardless of the type of circuitry, the preference decision was considerably affected by the amount of reverberation at the hands-free location; proximity operation was generally favored under conditions of moderate to high reverberation, nonproximity operation under conditions of low reverberation (high room constant). Other factors which affected the preference for hands-free proximity and distant talking are discussed.

I. INTRODUCTION

The hands-free telephone, which provides a convenient way of carrying on associated activities such as turning the pages of reference material, referring to drawings, etc., while the user is talking into a microphone and hearing from a loudspeaker located on his desk, has become an attractive supplement to the handset. This type of operation was first provided as a customer service in the mid-1950's in the form of the 595 telephone set, and soon thereafter in supplementary form as the No. 1A Speakerphone system,[1] neither of which employed voice switching.

The fact that under certain operating conditions hands-free operation can result in singing, in the transmission of reverberation or a barrel-like quality and in excessive noise for the handset listener at the other end of the line has been recognized. The effects of reverberation or liveness of the room and of talking distance have been appreciated. There

has been some speculation concerning the acceptability of reduced talking distance as a remedy for such operational characteristics.[2] More recently, switched gain, in which loss is introduced into the receiving path when the user is talking and into the transmitting path when the user is listening, has been explored as a remedy for some of these effects.[3]

This paper reviews the operational problems of hands-free telephony and gives the results of some experiments in which user reaction to, and acceptance of, various ways of achieving hands-free operation were studied. The tests were carried out in offices in which the "reverberation" was measured in one or more ways. Two different talking distances were provided, one by means of a microphone on an elevated arm about five inches from the lips (proximity talking) and a second by means of a microphone on the desk top about 20 inches from the lips (distant, or nonproximity, talking).

II. OPERATIONAL PROBLEMS OF HANDS-FREE TELEPHONY

To provide hands-free operation of the telephone, amplification must be introduced in both the transmitting (microphone) and receiving (loudspeaker) branches of the circuit. The amount of gain that can be so introduced in a properly installed set, before operational difficulties are encountered, is limited primarily by the acoustic properties of the location and by the hybrid balance afforded by the connecting line and trunk. These operational difficulties may be classified as follows:

(a) sustained feedback, or singing;

(b) enhanced sidetone, or return of the far-end subscriber's voice to him in the form of a reverberant echo;

(c) reduction in the transmitted signal-to-noise ratio and

(d) increased transmission of reverberant energy to the far end of the line.

2.1 *Sustained Feedback, or Singing*

If the loudspeaker of a hands-free set is placed too close to the microphone, or if the loudspeaker volume is turned up too high, the system will sing. The diagram of Fig. 1 indicates this will occur whenever the gain of the transmitting branch of the circuit, G_T, plus the gain in the receiving branch of the circuit, G_R, is greater than the loss of the hybrid coil, L_H, plus the loss of the air path, L_A. The transmitting gain is normally fixed for a given nominal talking distance in order to deliver a level to the central office that is comparable to that delivered by a

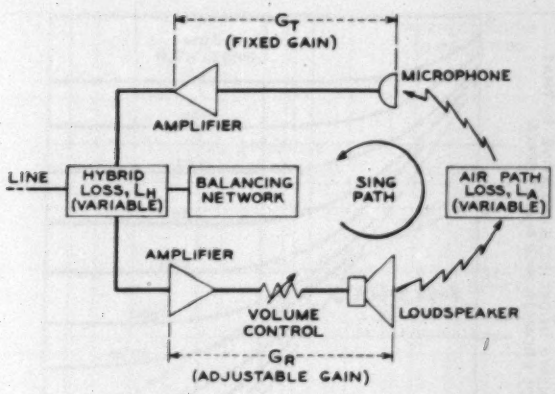


Fig. 1 — Electroacoustic path of signal for singing condition: $(G_R + G_T > L_H + L_A)$.

500-type set on a long loop. The gain in the receiving or loudspeaker branch is adjustable and under control of the hands-free subscriber. Singing will be initiated if this control is advanced too far, as it might be under noisy conditions, or if the incoming speech signals are low, or if the hearing of the hands-free subscriber is impaired.

Singing will also occur if either the hybrid coil loss or the air path loss become sufficiently small. The former depends on how well the impedance of the balancing network matches the impedance of the line. A varistor matching (balancing) network is used in order to provide some compensation based on variations in line resistance. The compensation is designed to be most effective for the medium-to-longer loop where higher gains are needed, rather than for the shorter loop connection where line losses are low.

The air path loss depends upon the distance between the loudspeaker and microphone and upon the relative amounts of sound which reach the microphone directly and by way of reflections from the walls, ceilings and furniture. If these surfaces reflect most of the energy which strikes them, the level at the microphone position falls off much more slowly as the separation is increased than it does when direct energy only is present.

Fig. 2 shows the ratio of the energy density $\bar{p}_r$ at various points within an enclosure to the energy density $\bar{p}_1$ at a distance of one foot from the same source in free space.[4] The curves are plotted for various values of the room constant R, which is defined by the relationship

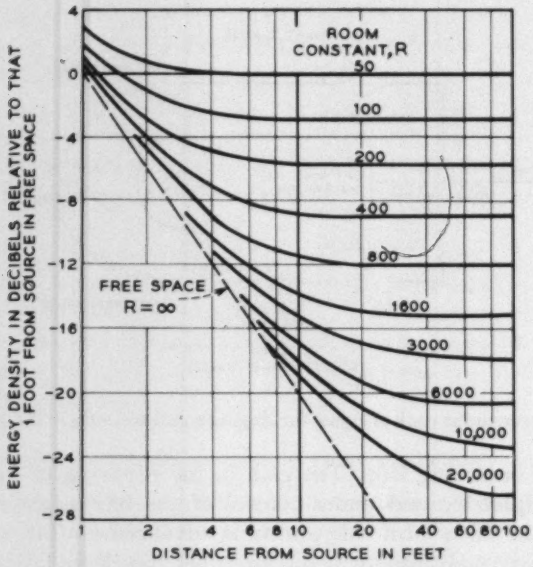


Fig. 2 — Average energy density vs. distance for various values of the room constant R.

$$R = \frac{\bar{\alpha}s}{1 - \bar{\alpha}} \tag{1}$$

$$\doteq \frac{0.05V}{t_{60}(1 - \bar{\alpha})} \tag{2}$$

to a first approximation. Here $\bar{\alpha}$ is the average absorption coefficient of the reflecting surfaces, s is the total area in square feet, V the volume in cubic feet, and t_{60} the reverberation time in seconds.* The value of R is large whenever the ratio of direct to reflected energy is also large (lower curves).

Fig. 2 shows that, in a very reverberant room, increasing the separation of the microphone and loudspeaker does not increase the air path loss very rapidly, although the direct unreflected sound ($R = \infty$) does decrease rapidly with distance, according to the inverse square law.

The next chart, Fig. 3, shows how the permissible (or usable) amplification is reduced as the acoustic environment departs from the ideal

* This is the time required for the sound to die away to one thousandth of its initial pressure, which corresponds to a drop in the sound pressure level of 60 db, following abrupt cessation of the generating source.

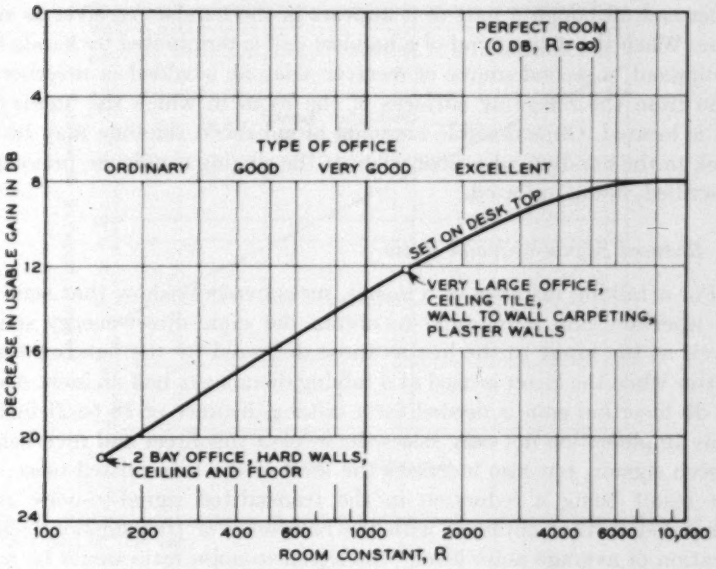


Fig. 3 — Decrease in usable gain as a function of the room constant R.

of free space.[2] Here, the abscissa is expressed in terms of R, the parameter of the previous family of curves. Fig. 3 shows that the introduction of a desk top or similar surface to support the hands-free set reduces the usable gain in an otherwise perfect room by about 8 db. A further loss results when the desk is surrounded by walls, ceiling, floor and office furniture. For an office with a considerable amount of acoustic treatment in the form of drapes, carpeting, acoustically treated ceiling, etc., R values as high as 2000 or more may be achieved. In such cases, the additional loss (beyond 8 db) may be held to a few decibels. In untreated offices, R values of 200 or less may result with an associated total loss in usable gain of 20 db or more. Fig. 3 shows that over the range of "ordinary" to "very good" offices, the room constant R varies from about 150 to 1500 and that about 9 db more gain may be employed in the latter compared to the former location.

2.2 *Enhanced Sidetone, or Far-End Talker-Echo*

When the *handset* subscriber talks into the transmitter, the principal component of the voice signal travels over the line to the party at the

other end, although a part of it appears in the handset receiver as sidetone. When the distant end of a handset call is terminated by hands-free equipment, a second source of receiver sidetone is added: a reverberant echo from the reflecting surfaces of the room in which the hands-free set is located. Objectionable amounts of enhanced sidetone may be fed back to the handset subscriber without the singing condition, previously described, being initiated.

2.3 *Reduced Signal-to-Noise Ratio*

For a talking distance of 5 inches, measurements show that some 13 db insertion gain is required to obtain the same direct-energy speech levels at the input to the line as those delivered by the handset transmitter when the latter is used at a talking distance of half an inch. About 25 db insertion gain is needed for a talking distance of 18 to 21 inches. This amplification not only raises the level of the direct and reverberant speech signals, but also increases the level of the transmitted noise, the net result being a reduction in the transmitted signal-to-noise ratio compared to that obtained with the handset. For transmission from a location of average noise level,* the signal-to-noise ratio would be quite acceptable for a 5-inch talking distance but would be rather noticeable for an 18- to 21-inch talking distance. For a noise ambient above about 60 db, the transmitted noise rapidly becomes very objectionable when a talking distance of 18 to 21 inches is used. For 5-inch proximity operation, this occurs for noise ambients above about 70 db.

Fortunately for the handset user in the case of a 5-inch talking distance, the hands-free user tends to revert to the handset before the noise transmitted to the handset end becomes intolerable; this is not so (because of the larger amplification of the transmitted signal) for an 18- to 21-inch talking distance.

2.4 *Transmitted Reverberation*

It has been recognized that the liveness or reverberant character of a room affects the ratio of direct-to-reflected energy which is transmitted to the far end of the line, and a curve has been given which relates talking distance and the room constant R at which reverberation effects are just below a noticeable level at the handset end of the line.[2] This curve is reproduced in Fig. 4, where it can be seen that R must be above 1500

* As used here, average noise refers to a level of 50 dbA (50 db above a reference acoustic pressure of 0.0002 dyne per cm^2, 40 db weighting). This value was reported by Inglis on the basis of several surveys covering a large number of installations.[5] Individual values, however, may differ from this value by a considerable amount even for similar locations.[6]

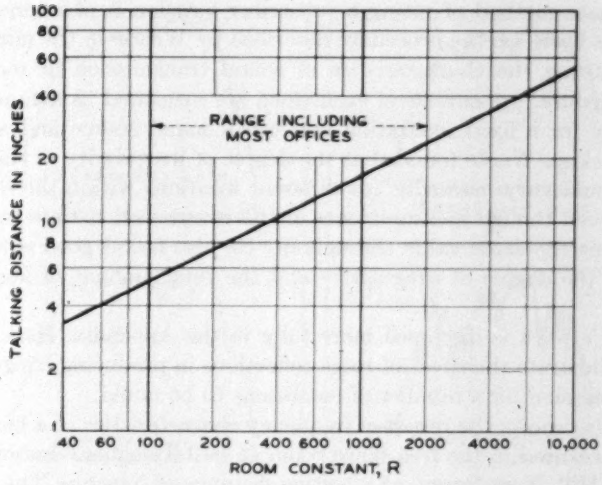


Fig. 4 — Maximum talking distance for no noticeable reverberation at the handset end of the line.

for a talking distance of 21 inches if reverberation effects are to be avoided, but may be as low as 90 for a talking distance of 5 inches. These values of R correspond to offices which are, respectively, well above and well below the average location from the standpoint of reverberation.

The room constant R is one of the parameters of the tests described in this paper. Unfortunately, this could introduce some error in the interpretation of the results. The room constant R is not a complete or completely reliable description of an acoustic environment, and, indeed, we are far from having any completely trustworthy assessment.

In particular, the room constant may fail to assess acoustic conditions of offices in which all of the acoustic treatment is concentrated on one or two surfaces and in which the installation places the microphone in the vicinity of highly reflecting surfaces. The value normally assigned to R on the basis of measurement is the average of several individual measurements except that "any one value that seems to be very inconsistent with the others (say, by a factor of two or more) should be disregarded, since this value probably results from an unusual condition in one section of the room."*

A knowledge of just such an "unusual" condition, however, is necessary if the individual subscriber's evaluation of the various methods of providing hands-free service is to be properly assessed. For this reason,

* From current instructions for making this type of measurement.

an alternate method of rating a subscriber location is of interest. This method is based on the procedure described by Wente[7] in the mid-1930's for measuring the characteristics of sound transmission in rooms. In this procedure, pressure-level excursions are measured as a function of frequency for a fixed separation between a sound source and a microphone pickup. Wente found that the degree of irregularity of the excursions "could vary markedly" at different locations within the same enclosure even though measurements of the reverberation time tended to give about the same value throughout. He also found good correlation between the degree of irregularity and the total amount of absorption present.

Wente's work is discussed more fully in the Appendix. Here we will merely illustrate the effect of room reflections in producing irregularities in transmission for a number of conditions to be noted.

Fig. 5(a) shows the response frequency characteristics of a hands-free set as measured in the free-space room at Bell Telephone Laboratories, Murray Hill, New Jersey, at a testing distance of 5 inches. The over-all smoothness of the curve indicates an essential absence of transmitted reverberation.

The response of Fig. 5(b) was obtained at the same testing distance of 5 inches, but in a rather reverberant room having an R value of about 150. This response is of interest since it represents excursions which are just below a noticeable level in the form of reverberant quality at the handset end of the line. Listening tests indicate this occurs whenever the ratio of direct to reflected energy is 10 db.

Fig. 5(c) shows the response obtained at the same testing distance of five inches in an intermediate class office having a measured R value of 805. A comparison with the response of Fig. 5(b) indicates that under these conditions the amount of transmitted reverberation would go unnoticed at the handset end of the line.

Fig. 5(d) shows the effect of increasing the testing distance to 21 inches at the Fig. 5(c) location. A comparison of Figs. 5(c) and 5(d) is of interest since they serve to illustrate the difference in the amount of reverberant content which is transmitted to the line for nominal 5-inch proximity and 21-inch nonproximity use, respectively. At more reverberant locations the quality contrast between these two talking distances would tend to be larger. Under improved conditions the difference would become increasingly less noticeable until, under free-space conditions, neither talking distance would result in the transmission of reverberant energy.

The response frequency characteristics of Figs. 5(e) and 5(f) are

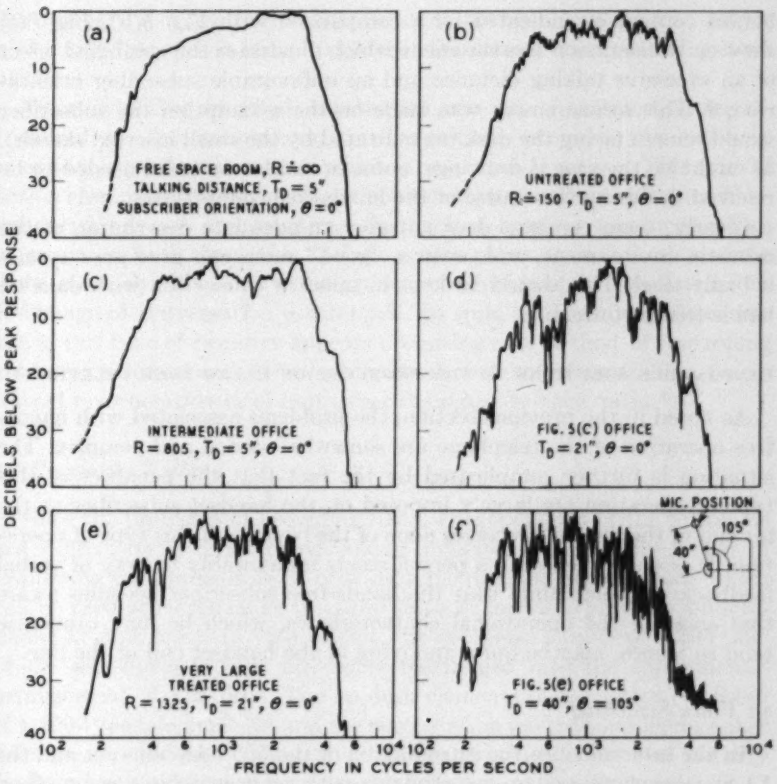


Fig. 5 — Response frequency characteristics of hands-free set for various conditions of operation.

examples of the transmission of excessive amounts of reverberant energy from a rather spacious office ($V = 5760$ cubic feet) which the large room constant indicates to be very good ($R = 1325$). Such a room constant should provide freedom from reverberation effects for talking distances up to 19.5 inches. In the present case, however, the set was installed on a side table in a corner of the office near highly reflective hard wall and window surfaces. This materially reduced the beneficial effects of the full ceiling treatment and the wall-to-wall carpeting. Fig. 5(e) shows a response frequency characteristic for a talking distance of 21 inches. Although this distance is only slightly in excess of the 19.5-inch separation noted above, the chart shows a rather noticeable amount or rever-

berant content as indicated by a comparison with Fig. 5(b). Fig. 5(f) shows a transmission measurement which illustrates the combined effect of an excessive talking distance and an unfavorable subscriber orientation, θ. This measurement was made on the assumption the subscriber would remain facing the desk (as indicated by the small inserted sketch), as might be the case if drawings, notes or similar material needed to be referred to during the course of the hands-free connection.

Clearly, room constant does not give an adequate description of the acoustic environment, and, even a "good" room can give poor results if badly used. This should be kept in mind in connection with data on hands-free performance.

III. POSSIBLE SOLUTIONS TO THE PROBLEMS OF HANDS-FREE TELEPHONY

As noted in the previous section, the problems associated with hands-free operation of the telephone are somewhat varied and complex. The situation is further complicated by the fact that the penalties of this type of operation are largely imposed on the handset subscriber at the far end of the line who receives none of the benefits of this type of operation. It is only when such a person reacts unfavorably by way of verbal feedback (or complaint) that the hands-free subscriber becomes aware that some of the operational characteristics, which he may otherwise tend to ignore, may be quite annoying at the handset end of the line.

3.1 *Voice Switching*

In the interval since the introduction of the 595 telephone set and the 1A Speakerphone system, experiments with voice switching as a method of improving hands-free operation of the telephone have been undertaken.* In such systems,[3] variolossers are normally employed in both the transmitting and receiving branches of the circuit in such a way that attenuation is introduced in only one circuit at a time. For one such system, the set is in the receiving condition during the quiescent or normal state. That is, incoming speech signals reach the loudspeaker without being attenuated by the variolosser in the receiving branch of the circuit. Outgoing signals, on the other hand, must switch out the attenuation of the transmitting circuit variolosser before reaching the line. The circuit returns to the quiescent or receiving state when either the hands-free subscriber stops talking or the level of the incoming speech is high enough to override the input from the hands-free microphone.

* The use of voice switching in communication systems is not new. One of its early uses was in connection with transatlantic two-way radio.[8]

An auxiliary rectifier circuit differentiates between loudspeaker output and hands-free speech input to prevent the former from switching itself off.

Voice switching, in which substantial amounts of attenuation are introduced by the variolossers, thus essentially provides two one-way telephone circuits, only one of which is activated at a time. This type of action eliminates two of the operational problems outlined in Section II — enhanced sidetone and sustained feedback of howl — since neither can occur unless both the active and the return paths are conducting at the same time. By this same action, however, a completely free-flowing interchange of conversation is inhibited. In spite of these limitations, the use of this type of circuitry appears promising as a method of controlling enhanced sidetone and sustained feedback, but not of controlling transmitted reverberation or of improving the signal-to-noise ratio.*

3.2 *Proximity Talking*

Since the problems of providing hands-free operation of the telephone are a direct result of increased talking distance, the most direct way of improving transmission is to provide a microphone arrangement which can be used at closer range than the 18 to 21 inches typical for a microphone at desk level. In view of this, the component parts of a 595 telephone set shown in Fig. 6(a) were rearranged into the experimental arrangement of Fig. 6(b). The on-desk elements of the latter consisted of a 500-type telephone set and the combination microphone-loudspeaker arrangement shown at the right. This experimental unit was initially used during a series of non-voice-switched trials at San Francisco.

To initiate a call with the arrangement of Fig. 6(b), the subscriber pulled the supporting arm forward. This rotated the microphone into the talking position shown and closed an ON-OFF switch located within the sphere at the base. Dialing was completed in the usual way.

The call could be transferred to the handset, if desired, by removing this unit from the cradle, as in ordinary use of the telephone. To transfer back to proximity operation, the handset was replaced on the cradle

* The use of a four-wire transmission line between calling stations also provides a means of eliminating all feedback paths except the round-trip path involving the acoustic coupling between the microphone and loudspeaker elements when two hands-free sets are used in a "back-to-back" connection. The chief disadvantages of this method are primarily those of cost and administration. A directive microphone might also be employed to some advantage for individual use. For conference application, however, directive restriction is not desirable. Also, the improvement from directivity decreases as the liveness increases, and thus it is least effective where most needed.

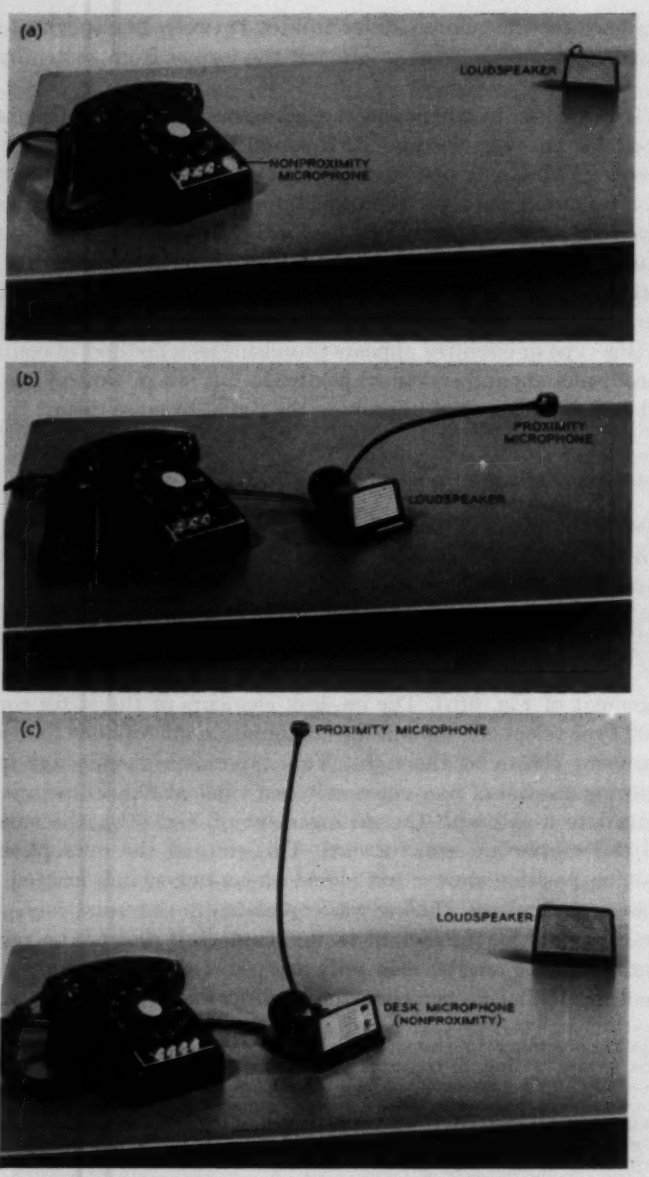


Fig. 6 — Hands-free set arrangements for proximity-nonproximity preference studies: (a) 595 telephone set; (b) proximity-talking microphone assembly; (c) combination proximity and nonproximity voice-switched telephone set.

while the proximity microphone was in the forward position. If it was replaced while the arm was vertical, the call was disconnected.

The location of the loudspeaker unit of Fig. 6(b) on the same base as the microphone had the advantage of directing the subscriber's attention in a common direction. It had the disadvantage of increasing the acoustic coupling between these two elements, thus decreasing the air path loss compared to that available with the arrangement of Fig. 6(a).

An alternate yet similar method which was used to provide proximity operation in a later experiment is shown by Fig. 6(c). Here a detached loudspeaker unit, similar to but larger than that shown in Fig. 6(a), was used. In this case the combination unit shown toward the center of Fig. 6(c) consisted of an arm-supported proximity microphone and a desk-supported nonproximity microphone as indicated. This arrangement was used during a second series of hands-free trials at Murray Hill office locations — this time in conjunction with the experimental voice-switched circuitry previously described.

Transfer from one microphone to the other was controlled by the position of the microphone arm. In the vertical position shown, the set operated as a nonproximity instrument; in the forward position, as a proximity device.

IV. PREFERENCE INDICATIONS FOR THE PROXIMITY AND NONPROXIMITY FEATURES

4.1 *The Non-Voice-Switched Trials*

This study was conducted in a San Francisco exchange area using 18 regular subscribers on a voluntary basis.* Since the original trial was intended to evaluate the performance of the 595 telephone set only, the installation of the proximity feature followed as a separate trial at stations where the 595 set had been in use for at least a month. Because of business travel and vacation schedules, the proximity sets remained installed at the various locations for a period of about eight weeks. Personal interviews were conducted at the conclusion of both series of tests. These interviews included several questions relating to customer reaction to the sets, their effectiveness under different operating conditions and their general acceptability in providing this type of service.

The subscriber locations at San Francisco varied from an office having a room constant of 399 to three having values above 1500. Eleven and

* The San Francisco trials were conducted under the immediate supervision of C. F. Benner of Bell Telephone Laboratories and D. S. Black of the Pacific Telephone and Telegraph Company.

possibly 12 were below an intermediate value of 1000. The great majority of the test participants were self-employed and/or dealing directly in customer relations — ten attorneys dealing with clients and court procedures, three insurance salesmen, one real estate salesman, one wholesale businessman, one person in the restaurant business and two for whom classifications are not available.

It would be expected that under these conditions proximity operation would find its greatest acceptance at locations where the room constant tended toward the lower end of the scale, and that nonproximity operation would find its greatest acceptance toward the higher end. The actual results, as summarized in Table I, show this to be the case. The preferences noted were obtained in answer to the final question of the interview series: "If a hands-free set similar to the one you have been using for the past few weeks were made available, which arrangement would you choose — that is, this one [Fig. 6(b)] or the set you had previously used [Fig. 6(a)]?" The results of Table I have been tabulated in order of increasing values of the room constant R.

The entries of Table I show that of 13 proximity votes ten came from locations having an R value below 1000, and that of five nonproximity (595) votes four came from locations having a room constant above 1000. Of the exceptions, the classification of one location was not obtained while the deviations of the other three from a rated value of 1000 were rather small. The significance of $R = 1000$ can be noted by again referring to the relationship of Fig. 4. For such a value, reverberation effects should go unnoticed at the hand-set end of the line for any talking distance which does not exceed 17 inches. As previously noted, this

TABLE I — PREFERENCE INDICATIONS FOR PROXIMITY AND NONPROXIMITY OPERATION UNDER SAN FRANCISCO TEST CONDITIONS

Subscriber Number	Preference	Room Constant, R	Subscriber Number	Preference	Room Constant, R
1	Prox.	low	10	Prox.	882
2	Prox.	—	11	Prox.	928
3	Prox.	399	12	595	947
4	Prox.	429	13	Prox.	1007
5	Prox.	474	14	Prox.	1061
6	Prox.	500	15	595	1235
7	Prox.	562	16	595	1566
8	Prox.	589	17	595	1633
9	Prox.	810	18	595	2412

assumes that the effective value of R at the position of installation does not differ significantly from the rated value.

4.2 *The Murray Hill Trials*

More recently, a second group of 18 individuals volunteered to participate in a comparison of proximity and nonproximity hands-free equipment — the circuits of both employing the voice-switched circuitry previously described. In addition, tests were also conducted without voice-switching, the two series requiring a combined total of about seven weeks testing time at each location. To insure independent preference judgments, the participants were instructed not to discuss any phase of the tests with other participants whom they knew or might later discover were also taking part in the study. In case contact by telephone with such individuals became necessary, they were requested to use the handset. For all other calls, they were asked to use the hands-free feature as much as reasonably possible.

During the voice-switched phase of the Murray Hill trial, the participants were provided with the combination proximity-nonproximity arrangement of Fig. 6(c). They were instructed to alternate between the two features from one call to the next and to transfer to the other feature during the call in case of comment from the far end of the line. Such a procedure provided for multiple direct comparisons of the two features and tended to eliminate practice effects in the use of generally unfamiliar equipment.* Approximately three weeks of testing time was allocated to this phase of the tests, which questioning indicated to be adequate. In no case did the participant feel that lengthening the test would have made any significant difference in his preference decisions.

The locations at Murray Hill covered an R-value range from 498 to 805 which, on this basis alone, would be expected to favor the selection of the proximity feature.

Columns 3 and 4 of Table II show the preferences voted on the basis of calls of a local nature only. They were obtained in answer to the following question: "Of the two types of hands-free service which you have been using, which would you prefer for your local calls if only one were made available? Allocate 100 points between the two to indicate your margin of preference." A tabulation of the higher of the two values only has been used to indicate both the direction and margin of the preference voted. The summation of the last row shows that of the 18 partici-

* One of the requirements for taking part in the trial was that the participant had made little, and preferably no, previous use of this type of equipment.

TABLE II — PREFERENCE INDICATIONS FOR PROXIMITY AND NONPROXIMITY OPERATION UNDER MURRAY HILL TEST CONDITIONS

Column: 1	2	3	4	5	6	7	8
Participant Number	R	Preference Decisions of Test Participants				Estimated Preference of Far-End Subscriber	
		Local		Long Distance			
		Prox.	Desk	Prox.	Desk	Prox.	Desk
1	498		60	60		75	
2	520	60		60		60	
3	538	75		95		100	
4	556	70		90		70	
5	588		90		60	100	
6	658		75		80	55	
7	663	60		90		70	
8	663	75		75		95	
9	673	60		65		80	
10	700		70	70		60	
11	706	85		100		75	
12	710	70		80		80	
13	710	80		100		80	
14	714	55		70		90	
15	720	65		75		90	
16	740		80		80		60
17	802		60		60	60	
18	805	75		60		65	
Votes for		12	6	14	4	17	1

pants, 12 voted in favor of the proximity feature for local call use compared to 6 votes in favor of the nonproximity or "desk"-type feature.

Columns 5 and 6 show the corresponding preference votes for use of the hands-free set for long distance calls, or calls for which the loop losses or incoming levels were noticeably poorer than those generally encountered on a local call. The summation of the last row shows 14 votes in favor of the proximity feature and four in favor of the nonproximity or desk-type feature under these conditions.

It may be noted that, in the above tests, the highest room constant was 805, and that in the San Francisco trial proximity talking was pre-

ferred by *all* users at locations having a room constant of 928 or less. Why did not all of the Murray Hill participants prefer proximity talking?

This might be attributed to the use of switched gain at Murray Hill. However, when the voice-switched sets were replaced by non-voice-switched equipment during approximately the latter half of the seven-week testing period, three votes were transferred from distant talking to proximity talking and four votes were transferred from proximity talking to distant talking, a net *gain* of one vote for distant talking for the use of non-voice-switched equipment. Since each participant had two votes (one for local calls and one for long distant calls) a transfer of one vote represents a net change of only one half of one full preference decision. During these trials, the proximity and nonproximity features were used during separate successive installation periods as had been the case for the non-voice-switched trials at San Francisco.*

Apparently there must be some explanation other than voice switching to account for the tendency of the participants at Murray Hill to use the distant talking feature under less favorable acoustic conditions. The writer believes that differences in motivation of the participants at the two locations and the increased use of the set for conference calls at Murray Hill largely account for the trend observed.

It has been noted that the participants in the San Francisco trial were chiefly self-employed men dealing directly in customer relations. It seems reasonable to expect that such individuals would be strongly influenced by complaints from the subscriber on the handset end of the line, and we have noted that it is only through such complaints that some of the defects of distant talking become apparent to the user of hands-free equipment.

In the Murray Hill trial, all participants were salaried employees and, as such, might not be expected to have the same amount of interest in achieving as good quality transmission to the far end of the line as would the participants at San Francisco. Thus, they could have had less reason for favoring proximity talking, even though they were aware that reception from the distant talking set was generally less acceptable to the handset subscriber than was reception from the proximity set. When asked: "What do you like best about the proximity feature?" their reply was essentially: "The reception was better at the other end of the line" or "There was less unfavorable comment." When the par-

* It should be noted that, because of changes in work location assignments, two of the 18 participants (Nos. 12 and 13) were unable to take part in this phase of the study. However, since both had previously voted twice for the proximity feature (Columns 3 and 5 of Table II), any change in their vote in the latter series of tests could only have further increased the gain noted above.

ticipants were asked: "What percentage of the subscribers on the other end of the line do you think would prefer that you use the nonproximity feature for your hands-free calls and what percentage the proximity feature?" the entries of columns 7 and 8 of Table II show that, in 17 out of 18 cases, the participants thought that a higher percentage of the far-end subscribers would prefer to be called by way of the proximity feature under the acoustic conditions which existed at the Murray Hill locations. Note, in particular, the ratings of participant No. 5, Table II.

Finally, when the six individuals in the Murray Hill trials who had voted for the distant-talking feature were asked to divide 100 points between the proximity and nonproximity features on the assumption they were now working on a fee or commission basis or were dealing directly in customer relations, all shifted their vote to the proximity feature to provide the 18 to 0 vote shown in columns 3 and 4 of Table III.* The crosses of column 3 correspond to participants whose columns 3 and 5 entries of Table II both show a proximity vote. It thus seems likely that participant motivation toward achieving favorable far-end subscriber reaction to the transmitted signal is a significant factor in the establishment of the preference decision of the hands-free subscriber.

In order to determine what was considered the most attractive features of the two methods of achieving hands-free operation of the telephone, the participants were asked: "What did you like best about the nonproximity feature? What did you like best about the proximity feature?" In answer to the former, the almost universal response was "convenience of use and/or more suited for conference connections." In answer to the latter, the reply, as previously noted, was essentially: "The reception was better at the other end of the line" or "There was less unfavorable comment."

The general absence of a 100 to 0 vote in favor of one or the other hands-free feature suggests that the participant might actually have preferred a set which would provide a choice of using either feature in order to more nearly meet the requirements of each individual call. Columns 5 and 6 of Table III show that, in response to such a question, 13 of the 18 participants indicated that this was so. Presumably each individual was also influenced in his rating by various calling needs.

The effect of the need for conference calls is illustrated by the data of columns 7 and 8. Here, column 7 gives the percentage of conference calls

* Here, one of the advantages of voting by division of points is illustrated. Under altered conditions, a re-evaluation of a similar appraisal is made possible without any implication on the part of the questioner that the participant should reverse his previous vote. For example, any one participant might simply have altered a previous division of points without effecting a reversal.

TABLE III — PREFERENCE INDICATIONS FOR PROXIMITY AND NONPROXIMITY OPERATION UNDER MURRAY HILL TEST CONDITIONS

Column: 1	2	3	4	5	6	7	8
Participant Number	*R*	Vote With "San Francisco" Motivation		Features Desired By Participants		Estimated Percent of Conference Calls	Votes For Desk Feature From Table II
		Prox.	Desk	Both	1st Choice Only		
1	498	90		75		25	1
2	520	×		60		0	
3	538	×		70		0	
4	556	×		70		5	
5	588	75		70		10	2
6	658	90			100	5	2
7	663	×		60		0	
8	663	×		80		5	
9	673	×		75		0	
10	700	70			60	40	1
11	706	×			85	0	
12	710	×			60	1	
13	710	×		60		0	
14	714	×		80		0	
15	720	×		100		2	
16	740	60		75	100	0	2
17	802	70		65		5	2
18	805	×				15	
Votes for		18	0	13	5	Average: 6.3	

as estimated by the individual participants. In general, those making this type of call showed a greater preference for the nonproximity feature than did those making only individual calls. Column 8 shows that eight of the ten votes for the nonproximity feature were from conference call participants. And, since the estimated average percentage of conference calls at Murray Hill was more than five times as large as the average 1.2 per cent reported at San Francisco, it is likely the greater use of this type of service at Murray Hill was also a factor toward the increased preference for the nonproximity feature at the latter location.

V. OTHER FACTORS AFFECTING THE PREFERENCE DECISION

In addition to the various factors which have been shown to affect the selection of one hands-free feature over the other, there remain the differential effects of such factors as microphone response, physical design and the amount of switched loss employed, which for the present tests was quite large. In terms of microphone response, the output of each unit was equalized to give approximately the same response over a frequency range of about 350 to 3500 cps when used at their respective nominal talking distances. For this reason, no attempt has been made to apply any correction to the preference appraisals on this basis. Nor has any correction been attempted on the basis of physical design, since the tests were limited to but one of several arrangements that might have been chosen to provide the proximity feature.

In connection with the amount of voice-switched loss employed, it should be noted that during the Murray Hill trials no reduction was made in this quantity when the proximity microphone was in use, although the lower gain associated with this type of operation would have permitted such an adjustment. This procedure was followed because the circuit to which the proximity feature had been appended did not lend itself to such an interfeature switchover adjustment. Subsequent testing, however, indicated that such a reduction would have resulted in improved proximity operation.

VI. SUMMARY

As the acoustic environment at the hands-free location departs from the ideal of highly absorbent surrounding surfaces and a low ambient noise level, the basic problems of furnishing this type of service become increasingly more difficult, particularly at talking distances comparable to those used during conference-type connections. The problems of enhanced sidetone and sustained feedback can be adequately controlled by the use of voice-switched circuitry, but not without the introduction of other operational difficulties. An improved form of voice switching on an adjustable basis and with careful attention to transient performance holds promise of reducing these effects to rather acceptable levels. This improved operation, however, still leaves unsolved the remaining problems of reduced signal-to-noise ratio and the reverberant quality of the transmitted signal.

In both of two trials involving proximity and distant talking, the proximity feature was preferred over distant talking. In the San Francisco trial, proximity talking was preferred in all instances in which the

room constant was less than 947. In the Murray Hill trial, the highest room constant was 805, yet distant talking was preferred by 5 out of 18 participants under these conditions.

The use of voice switching in the Murray Hill trial might seem to be an explanation of this difference. However, removal of the voice-switching feature affected the preference for distant talking by an amount not judged to be significant. The fact that the room constant is not a completely adequate measure of acoustic environment might account for some of the disparity. It is the writer's belief, however, that the factors of greatest significance are: First, the participants at Murray Hill as salaried employees were less strongly motivated to please the person with whom they talked than were the participants at San Francisco, who were largely self-employed, and second, there was greater use of the hands-free set for conference calls at Murray Hill. Replies of the Murray Hill participants to questioning are in conformity with the first of these concepts, and the greater preference shown at Murray Hill for the distant-talking feature by conference-call participants is in conformity with the second.

The results at Murray Hill show that a majority of the participants at this location indicated a preference for having both the proximity and distant-talking features rather than either alone.

VII. ACKNOWLEDGMENTS

In addition to acknowledgment given by previous reference, the author wishes to express appreciation to the several members of the Apparatus Development and Transmission Engineering Departments who were associated with the present study in one way or another. He also wishes to thank the 18 members of staff who participated in the Murray Hill trials. Finally, the author wishes to express a particular appreciation to John R. Pierce, Director of Research and Communication Principles for his encouragement and support, and to E. E. David, H. L. Barney and other members of the Visual and Acoustics Research Department for similar support and assistance during the course of the present study.

APPENDIX

Transmission Quality Based on Frequency Response of Path Between Talker and Microphone

Wente[7] defined the degree of irregularity of the sound pressure level excursions of a response frequency characteristic as the sum of the

pressures of all of the maximum points minus the sum of the pressures of all of the minimum points over the frequency range included, i.e.,

$$\text{D.I.} \equiv \sum_{f_1}^{f_2} p_{\max} - \sum_{f_1}^{f_2} p_{\min} \qquad \text{(in arbitrary units).} \tag{3}$$

Alternate definitions of the degree of irregularity have also been proposed since the appearance of Wente's original paper by Bolt and Roop[9] and by Schroeder,[10] the latter having defined the irregularity in terms of the average excursion $\bar{h}$ of successive fluctuations, i.e.,

$$\bar{h} \equiv \frac{\sum_{f_1}^{f_2} (10 \log_{10} p_{\max}^2 - 10 \log_{10} p_{\min}^2)}{N} \text{ decibels,} \tag{4}$$

where N is the number of pressure peaks ($p_{\max}$) or pressure valleys ($p_{\min}$) between the frequencies f_1 and f_2 over which the summation is taken.

While the exact relation between irregularity, as defined, and the ratio of the direct to reverberant energy at the microphone position is an unsolved problem, measurements of this kind provide a direct means of indicating the effect on signal transmission of such parameters as talking distance, microphone placement and subscriber orientation, which a measurement of the reverberation time t_{60}, or a measurement of the room constant R as outlined in Section 2.4, does not give.

In investigating environmental effects by such a measurement at a hands-free location, a sound source having approximately the directivity of the human mouth and head was driven by a sweep frequency oscillator whose driving mechanism was coupled to an X-Y recorder. The sound output of the source was picked up by the microphone of the hands-free set whose output, in turn, was connected to the signal input of the recorder.* In order to simulate actual conditions of use, the sound source was placed at a position in space corresponding to the location which a subscriber's head would assume during actual use of the set.

When the distance from the source is small for such a measurement, the energy which arrives directly at the microphone position may almost completely override the reflected energy, and the response will be smooth. At large distances, on the other hand, the relative amount of reflected energy will be appreciable. At some frequencies the net reflection will

* In the present case, particularly for very large pressure excursions, the frequency sweep rate was not slow enough to permit steady-state conditions to be fully established at the microphone position. In actual practice, the trace obtained also depends on the writing speed of the recorder, its frequency bandwidth and the degree of irregularity introduced by variations in the sensitivity or efficiency of the test equipment.

be in phase with the direct sound, giving a peak. At other frequencies the net reflection will be out of phase with the direct sound, giving a valley. Thus a measurement made at a large distance in a highly reflective room or at an unfavorable subscriber orientation or microphone position within the room will show large fluctuations corresponding to the transmission of a large amount of reverberant energy.

The effect which changes in the above parameters have on the amount of transmitted reverberation has been illustrated by the various cross comparisons of the response frequency characteristics of Fig. 5 in Section 2.4 of the text.

REFERENCES

1. Clemency, W. F., Romanow, F. F. and Rose, A. F., The Bell System Speakerphone, A.I.E.E. Trans., **76**, Part I, (Comm. and Elect. No. 30), 1957, p. 148.
2. Emling, J. W., General Aspects of Hands-Free Telephony, A.I.E.E. Trans., **76**, Part I, (Comm. and Elect. No. 30), 1957, p. 201.
3. Busala, A., Fundamental Considerations in the Design of the Voice-Switched Speakerphone, B.S.T.J., **29**, 1960, p. 265.
4. Hopkins, H. F. and Stryker, N. R., A Proposed Loudness-Efficiency Rating for Loudspeakers and the Determination of System Power Requirements for Enclosures, Proc. I.R.E., **36**, 1948, p. 315.
5. Inglis, A. H., Transmission Features of the New Telephone Sets, B.S.T.J., **17**, 1938, p. 358.
6. Knudsen, V. O. and Harris, C. M., *Acoustic Designing in Architecture*, John Wiley and Sons, New York, 1950.
7. Wente, E. C., The Characteristics of Sound Transmission in Rooms, J.A.S.A., **7**, 1935, p. 123.
8. Wright, S. B. and Mitchell, D., Two-Way Radio Telephone Circuits, B.S.T.J., **11**, July 1932, p. 368.
9. Bolt, R. H. and Roop, R. W., Frequency Response Fluctuations in Rooms, J.A.S.A., **22**, 1950, p. 280.
10. Schroeder, M. R., Die statistischem Parameter der Frequenzkurven von grossen Raumen, Acoustica, **4**, 1954, p. 594.

Magnetic Latching Relays Using Glass-Sealed Contacts

By P. HUSTA and G. E. PERREAULT

(Manuscript received May 17, 1960)

A new type of relay making use of sealed contacts in association with one or more permanent magnets has been developed. The relay is operated and released by selective polar energization of the winding, and remains in the operated or released state following termination of winding energization. The design affords high speed, good sensitivity, circuit flexibility and power economies in various applications.

I. INTRODUCTION

Latching relays or lock-up relays, as they are sometimes called, are becoming important and very useful as switching elements. As a broad definition, a latching or lock-up relay is one whose contacts will remain in the operated state by means of mechanical or magnetic locking or because of reduced energy furnished to the coil or to a separate winding of the coil through a set of auxiliary contacts.

Latching relays take several forms, and some similarities and differences in the characteristics of this family of relays will be described. These relays are usually polar. Their most important characteristic is that they are sensitive, and therefore the relays are usually used where power is at a premium or must be conserved. Since they normally are pulse-operated, much power can be saved when a long period of operation of the relay in its operated state is required.

Many designs of typical armature-operated relays fall into the latching or locking classification of relay. This paper, however, will deal only with magnetic latching types of relays using glass-sealed contacts.

II. SEALED MERCURY CONTACT RELAYS

Although the earliest relays which contained glass-sealed contacts were of the dry reed type, the first sealed-contact relays to go into large scale production in the Bell System were those containing sealed con-

tacts of the mercury-wetted types. These sealed contacts, illustrated in simplified form in Fig. 1, provide a transfer by contacts wetted with mercury. When associated with a coil, magnetic structure and housing, they constitute one form of the well-known Bell System mercury-type relay. If this same structure also includes an adjustable permanent magnet, as shown in Fig. 2, then polar-biased and polar-latching relays result.

The various operating characteristics of this family of relays are shown by a series of staff diagrams in Fig. 3. Here all the function points of these relays are plotted in ampere turns. The crosses are points where the front contacts close or operate, and the circles are points where these contacts open or release. The crosses (X_1 to X_1) also show the variation in sensitivity of all the product, and the circles (O_1 to O_1) show the variation in release of all the product. For an individual relay, operate and non-operate coincide, as do hold and release. The relay shown in staff 1 has no magnet and is a neutral relay since it will operate at the same value of ampere turns for either polarity of coil current, while all the other relays shown are magnet-biased relays. Staffs 2, 3, 4 and 5 are biased polar relays. They are not true polar relays, because they have two sets of function points, the primary points X_1 and O_1 and the secondary points X_2 and O_2. These relays are called polar, however, because the desired function points are obtained only when the correct polarity

Fig. 1 — Sealed mercury-wetted contacts.

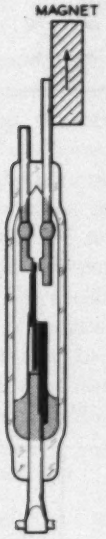


Fig. 2 — Magnet-biased sealed mercury-wetted contacts.

is applied to the relay. All secondary function points of biased polar relays are usually undesirable from a circuit standpoint.

There is one case where a reversal in selection of function points is of advantage. Where a high ratio of release to operate is required, the secondary function points are selected for the wanted operating characteristics of the relay, and the primary function points are avoided by selecting the opposite polarity in the use of the relay.

Since all of these magnetic-biased relays have only one capsule of sealed contacts, it is possible by adjustment of the biasing magnet to reduce the variation in operate X_1 points or the variation in release O_1 points. Practically, the adjusting error allows a reduction in spread of operate or release to about ± 5 ampere turns. There thus can be obtained in these relays an adjustment of weak or stiff-controlled operate or weak or stiff-controlled release. Staff 2 shows a relay adjusted to close operate with a weak adjustment. Staff 3 shows a relay adjusted to close operate with a stiff adjustment. Staff 4 shows close release with weak adjustment and staff 5 shows close release with stiff adjustment.

When the magnet is adjusted to cause the hold and release primary function points to occur on negative ampere turns, then we obtain the

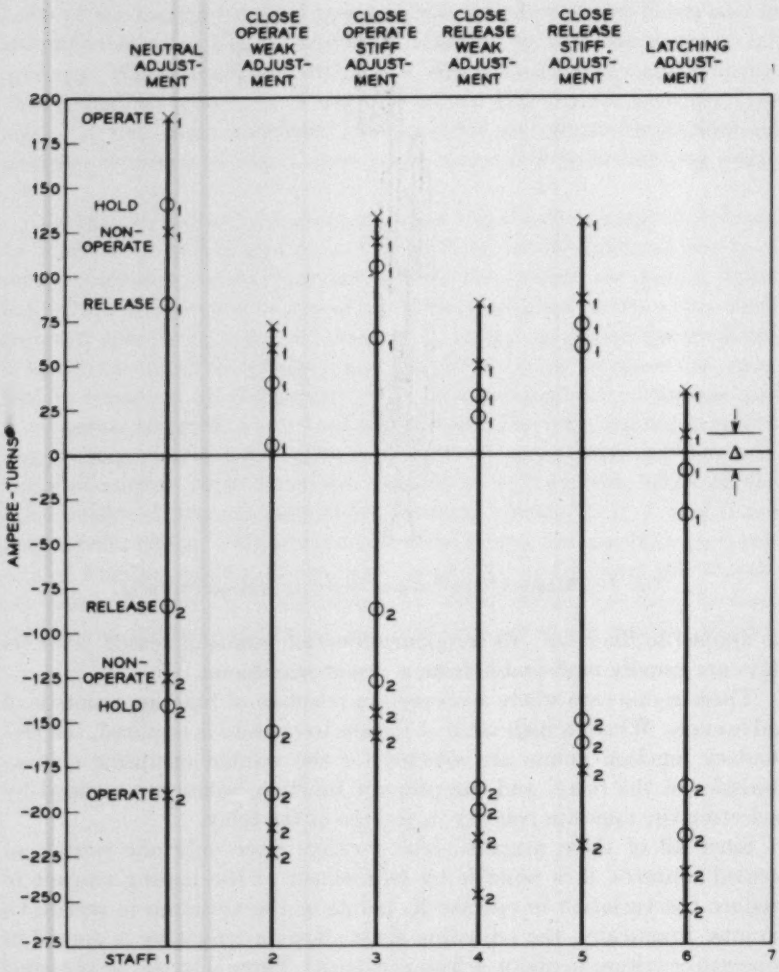


Fig. 3— Static sensitivity characteristics.

latching relay shown by staff 6. This relay will operate on positive pulses, will stay operated or latched when the power is removed and will release on negative pulses and stay released when the power is again removed. In order to be able to adjust a relay to this type of adjustment, the nonoperate function point must be higher than the hold function point and there must exist a margin or delta (Δ) between nonoperate

and hold, as is seen to be present with the magnet-biased relays in staffs 2, 3, 4, 5 and 6.

It will be noted, on re-examination of staff 1, that the variation of the operate values of a group of neutral relays overlaps the variation of the release values. Taken as a group, the neutral relays cannot make multiple sealed-contact latching relays because there does not exist a delta or positive difference between the nonoperate and hold function points. If individual magnetic adjustment to each capsule of sealed contacts can be made, or if sealed contacts are selected so as to group them to nearly the same sensitivity, then it becomes possible to make multiple sealed-contact latching relays. The sensitivity of magnetic latching relays of the same typical design of magnetic structure, coil, etc. as illustrated by the relays in Fig. 3 is two to six times better than that of neutral and biased relays of the same family, and since they operate on pulses they permit conservation of holding power.

III. BALANCED POLAR MERCURY RELAYS

There is another family of relays of great interest which uses glass-sealed mercury-wetted contacts. These are of quite a different design, and are even more versatile and sensitive. In this type of relay the sealed contacts are controlled by the precise adjustment of flux from two separately adjusted magnets[1] located on the pole pieces of the sealed contacts, as shown in Fig. 4.

With proper control of the adjustment of these magnets, this type of relay can provide a wide choice of operate values and a wide choice of release values. These choices of adjustments are shown in Fig. 5.

These relays may be adjusted to operate at any value shown on the

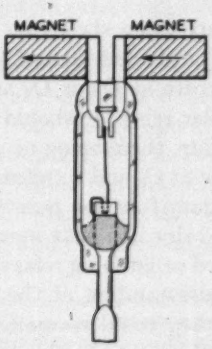


Fig. 4 — Magnet-controlled sealed mercury-wetted contacts.

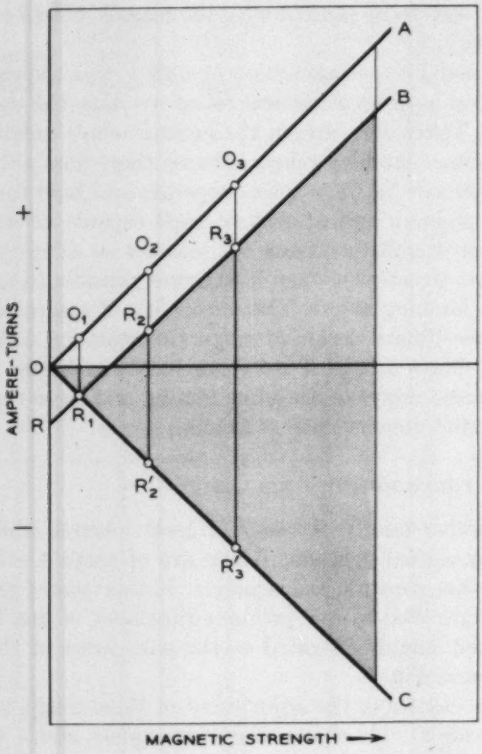


Fig. 5 — Magnet-controlled sealed-contact relay adjustment capability.

line OA and to release at any value shown on a staff directly below the selected operate value in the shaded area of the diagram.

A relay adjusted to operate at O_2 or O_3 and to release at R_2 or R_3 respectively is a biased polar relay. It should be noted that it has only these primary function points, there being no secondary function points. A relay adjusted to operate at O_1 and to release at R_1 is a polar latching relay. It also has no secondary function points.

If similar types of sealed dry contacts were to be manufactured, the same kind of neutral, biased or latching relays could be made. However, a complete study and understanding of the behavior of dry contacts would be required before they could be used. This is necessary because sealed mercury-wetted contacts obviously do not require high contact forces in order to produce good electrical connections, whereas sealed dry

contacts do. Some of the questions which must be answered are: What effect do these various adjustments have on contact force and chatter? Can adequate margins be obtained to assure reliable contact performance under both normal conditions and environmental conditions such as shock and vibration?

IV. SEALED DRY CONTACT RELAYS

The invention and development work on sealed dry contacts, which preceded the sealed mercury-wetted contacts, were the basis of many experimental and many paper relay designs. Of necessity, in order to find large-scale application, such sealed contacts must be inexpensive to manufacture, and thus their construction must be suitable for automation. Such a sealed contact (Fig. 6) was developed and it was shown[2] in 1955 that, in addition to a neutral relay with this simple sealed make contact, magnetic latching relays with makes, breaks or transfers could be produced. It was also shown that the sealed dry contacts would be functionally suited for all types of multicontact relays.

With multicontact relays, the spread of the function points of the sealed contact product must be taken into account before the design of magnetic latching relays can be attempted. The distribution of the function points of the dry sealed contacts is shown in staff 1 of Fig. 7. It will be noted that this distribution appears to be ideally suited for magnetic latching relays because there exists a large delta or spread between the nonoperate and hold functions. This spread makes it possible for these sealed contacts to be utilized without selection or separate adjustment, and also for some of the available margins to be used to absorb variables encountered in adjusting more than one enclosed contact with a single biasing magnet, thus simplifying the design of the latching multicontact relay.

When adjusted, a magnetic latching relay must have good margins against nonoperate and hold conditions at zero coil flux. This is to prevent the sealed contacts from falsely operating or releasing when the relay is subjected to vibration and shock. Because of the normal differences in sensitivity of the sealed contacts, and because of the unequal distribution of flux from the magnet to the sealed contacts, these relays must be carefully designed and precisely adjusted.

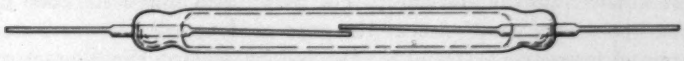


Fig. 6 — Sealed dry contacts.

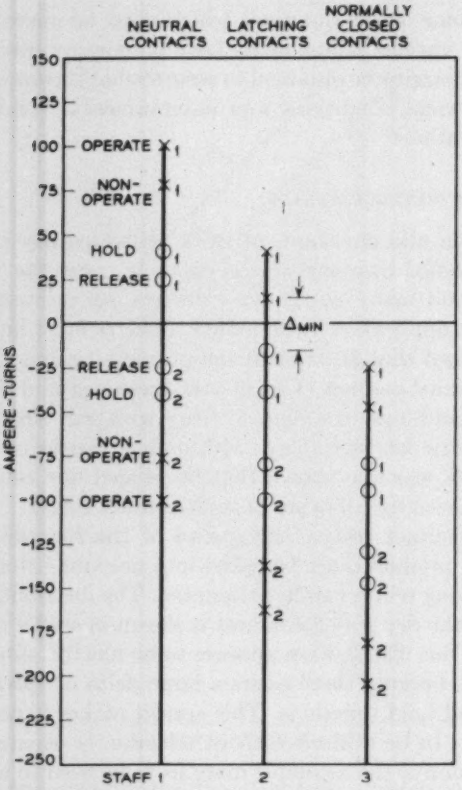


Fig. 7 — Typical characteristics of neutral and magnetic-biased relay.

Another important consideration is to obtain a design in which the secondary function points of Fig. 7 are, if possible, spread from the primary operating function points. In this way reclosure will be prevented when the release current is maximum in any given circuit. It has been found experimentally that the secondary function points can be spread by proper design of the permanent magnet and the magnetic return paths. A typical design is seen in Fig. 8. Here, four sealed contacts are grouped together so that, the glass envelopes being cylindrical, a space exists in the center of the group. The permanent magnet is then placed in this space at a position near the gaps of the reeds. A near-maximum volume of magnetic material is achieved by making the magnet square in cross section.

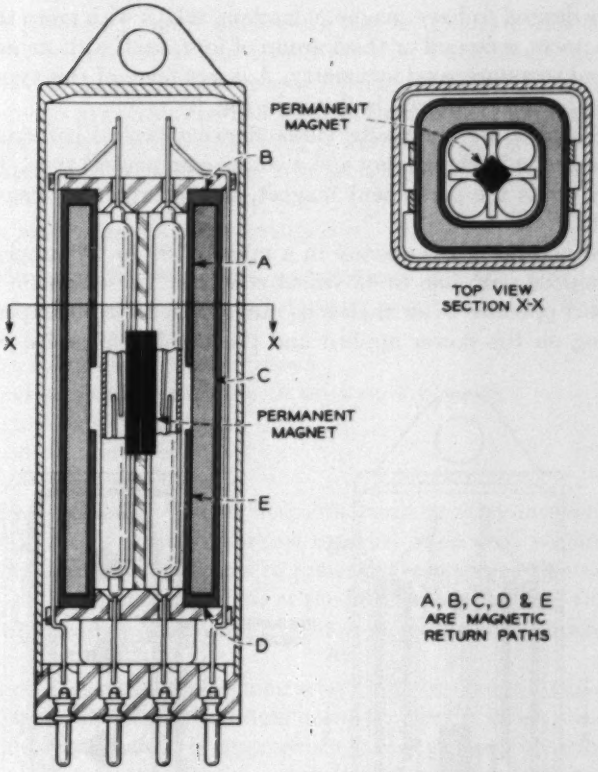


Fig. 8 — Typical four-contact relay design.

The precise adjustment of the magnet is accomplished by magnetizing it to saturation and then demagnetizing it while it is in place in the relay. The finished relay is adjusted by means of a servo-type adjusting set, which first magnetizes the magnet fully and then automatically demagnetizes it in small steps until all sealed contacts respond to the predetermined sensitivity selected for the relay and checked for by the servo-adjusting machine.

This basic design, incorporating one to four sealed contacts, will provide a relay with one, two, three or four contacts of either the neutral type or the magnetic latching type in the same size and same over-all structure, and with a minimum of parts. A single magnet might also be used in this manner for controlling more than four sealed contacts, but,

when it is desired to have magnetic latching relays with more than four sealed contacts, a second or third group of four, each with its own magnet, is used to assure good symmetry. A larger relay of this type, incorporating 12 sealed contacts, is shown in Fig. 9.

Figs. 8 and 9 are composite views showing typical construction of both a magnetic latching relay and a companion neutral type. The neutral relay omits the permanent magnet, but incorporates magnetic return paths.

Typical neutral relays operate in a range of 65 to 120 ampere turns when equipped with one to 12 sealed contacts. The magnetic latching counterpart operates in an analogous range from 45 to 90 ampere turns. Depending on the power applied and the number of sealed contacts,

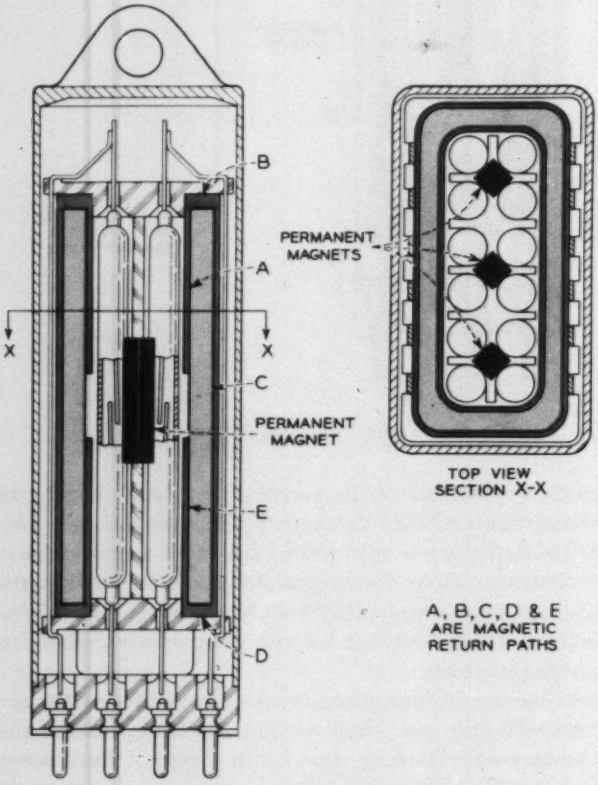


Fig. 9 — Twelve-contact relay design.

speeds of between 2 and 5 milliseconds can be obtained with the one to four sealed contact relays and speeds between 4 and 10 milliseconds on the larger 12 sealed-contact relays. Operating times of the magnetic latching relays are about one-half those of the neutral relays for the same power. Since they require no holding power, they may be pulse operated and released, as discussed in the next sections.

V. RELAY PERFORMANCE CHARACTERISTICS

In addition to flexibility, compactness and low unit cost, the relay characteristics sought by a circuit designer usually fall into the following pattern:

(a) fast operate and release times;
(b) good sensitivity or ability to work on low power;
(c) high reliability and long life;
(d) good contact force;
(e) low contact resistance.

The glass-sealed dry contact magnetic latching relay meets all of the above objectives with good working margins when used within its prescribed limits. A review will now be presented of the relay's performance, together with an analysis of some of the problems peculiar to this design of relay that require special consideration regarding the application and use.

Sealed-contact magnetic latching relays may be used with all contacts normally open or with all contacts normally closed. When the winding is energized by a current of appropriate polarity, magnitude and duration, the open contacts will close or the closed contacts will open. Upon removal of the winding energization, the contacts of these relays remain in their closed or open states respectively. In the usual circuit application of conventional neutral relays the expression "operate" generally refers to energizing a relay to cause its open contacts to close or its closed contacts to open and "release" refers to de-energizing the winding to cause the contacts to restore to their normal state. Since "operate" and "release" of the magnetic latching relays are both accomplished by energizing the relay winding with current of proper polarity, the customary reference to operate and release tends to become confusing if applied to describing the dynamic characteristics associated with the closing of open contacts and the opening of closed contacts. For the sake of clarity, therefore, the following discussion will be concerned only with relays intended for use with the contacts normally open, so that "operate" will refer to the act of closing the contacts and "release" to the opening of the contacts.

5.1 *Operate Case*

When the winding is energized to operate the relay, sufficient flux must be produced to cause the reeds to come together. With relays having only one sealed contact, there is no problem, since all of the coil flux is available to attract this single pair of reeds together. However, when more than one sealed contact is to be operated by the coil flux, we have a somewhat different situation. If the coil current in a multiple-contact relay is gradually increased, the contacts will close sequentially, depending on the relative contact sensitivities and the distribution of the available flux. It is necessary therefore to raise the coil current enough to produce sufficient flux to close the last remaining open contact. In ac-

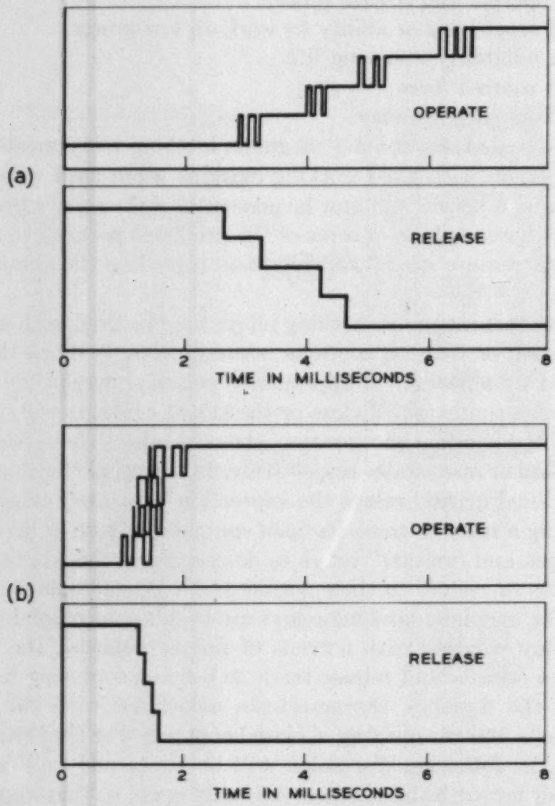


Fig. 10 — Typical sequential action of the four contacts in a magnetic latching relay: (a) low power energization; (b) high power energization.

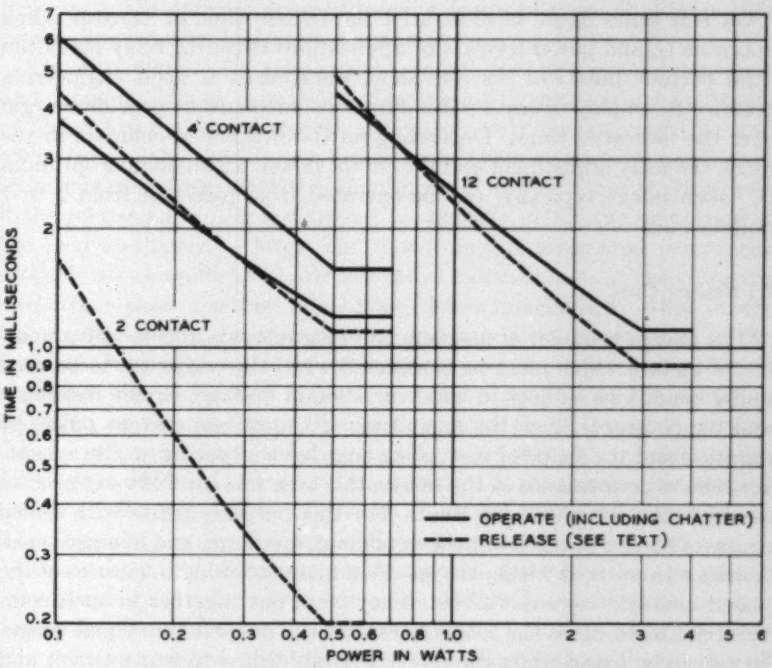


Fig. 11 — Relay function times at optimum G_c .

tual operational use and with suitably designed relays to meet short operate times, the rate of flux rise is sufficient to impart the necessary acceleration to all the reeds, and this minimizes but may not entirely eliminate the sequential contact closing in a multiple contact relay. This action is illustrated in Fig. 10: When the contacts close they chatter briefly.

The operate time as defined in this paper is the elapsed time from the start of coil energization to the closure of the last contact, to the end of chatter. It is a function of the coil constant G_c* and the total coil circuit power. Where short operate times are desired, the relay design aim is to optimize G_c for the particular circuit and power requirements. Fig. 11 illustrates the range of operate times measured to the end of chatter, for a 2-, a 4- and a 12-contact magnetic latching relay at various power levels and corresponding optimum G_c.

* $G_c = N^2/R$, where N is the number of turns in winding and R is the total coil and series circuit resistance.[3]

On this same figure is illustrated the release time at corresponding optimum G_c and power levels. For applications requiring relay operation from current pulses of discrete short duration it is good engineering practice to employ operate pulse durations adequate to provide margin over the indicated times. Depending on the number of contacts in the relay, the relay adjustment and the circuit power, and assuming optimum G_c, such relays typically, can be operated from pulses of from 2 to 7 milliseconds.

5.2 *Release Case*

The release behavior of magnetic latching relays is affected by a number of factors which must be considered when the relays are to be used under conditions subject to wide variation of voltage, circuit resistance and temperature. Since the magnitude of the release current pulse, its duration and the shape of its trailing edge have a bearing on the consistent release performance of the relays, this area was carefully explored to establish the relay working limits. For this purpose, relays with sealed contacts chosen to represent the maximum, minimum and average sensitivities encountered within the specified manufacturing tolerances of the sealed contacts were assembled. They were put together in such combinations as to cause the relays to exhibit the possible functional irregularities to be found when the relays were subjected to worst circuit and ambient conditions.

In Fig. 7, staff 2, it is shown that the relays can be caused to operate under two conditions of coil energization. In the first, or normal, case, the polarity of the coil current is such that the permanent magnet flux and coil flux are additive. When the relay is released the polarity of the coil current produces a flux which ideally should just cancel the permanent magnet flux, thus allowing the contacts to open. Under a variety of circuit applications, the release current could possibly be appreciably higher in magnitude than the operate current, though of opposite polarity, and the relay would release satisfactorily. However, if the associated circuit conditions are such that the release current can be considerably higher than normal, the release coil flux may then be so high that it will not only cancel the permanent magnet flux but the net difference flux could be high enough to hold one or more contacts closed. In this case, although there is a flux reversal in the reeds the reversal occurs so rapidly that the contacts do not open. This is the secondary operate point illustrated in Fig. 7, and its possibility must be considered if the relays require sensitive adjustments and are likely to be applied under conditions involving wide variations in voltage, variations in external circuit

resistance, low ambient temperatures or fortuitous circuit trouble that might produce very high release current.

The application of a release pulse causes the reed contacts to open and, being undamped cantilevers, they vibrate at their natural frequency of about 840 cycles per second, alternately moving close together and away from each other during the decay of their mechanical oscillation, as long as the release current is maintained in the coil winding. However, if the release current is terminated too quickly, while the amplitude of the reed oscillation is large, the flux from the associated permanent magnets may be sufficient to reclose one or more contacts as the vibrating reeds come close together. Some may reclose momentarily while others may reclose permanently, depending on the relative sensitivities of the contacts. Normally, when the reeds are at rest the gap reluctance is so high that the permanent magnet cannot by itself close the contacts. However, on the small gaps that occur alternately as the reeds vibrate, the gap reluctance may be low enough to cause contact reclosures under limiting conditions. During one time constant of the damped mechanical oscillation decrement there is a region of critical points in time at which these irregularities could occur, as illustrated in Fig. 12.

Consistently reliable release performance is obtained when the release pulse is maintained for a time sufficient to allow the mechanical oscillation to drop to safe limits. This time (typically 5 to 30 milliseconds) will vary from one relay application to another, depending on the number of contacts in the relay, its adjustment, the circuit power and the character of any electrical paths in shunt with the relay coil circuit. The latter may affect the trailing edge of the release pulse. If the release current decay is oscillatory the fortuitous phasing of the mechanical reed oscillation and the reverse current transient may enhance reclosures. Steep current decays cause the contacts to come immediately under the influence of the full permanent magnet flux during the time of large reed oscillation, when the reeds can come close to each other. By choosing proper values of resistance and capacitance connected in series and bridged across the contacts that de-energize the relay, this effect may be minimized in a given application by effectively damping the coil current transient.

5.3 *Contact Characteristics*

5.3.1 *Contact Force*

Consistent and reliable performance of any relay demands that when its contacts are closed they must press together with sufficient force

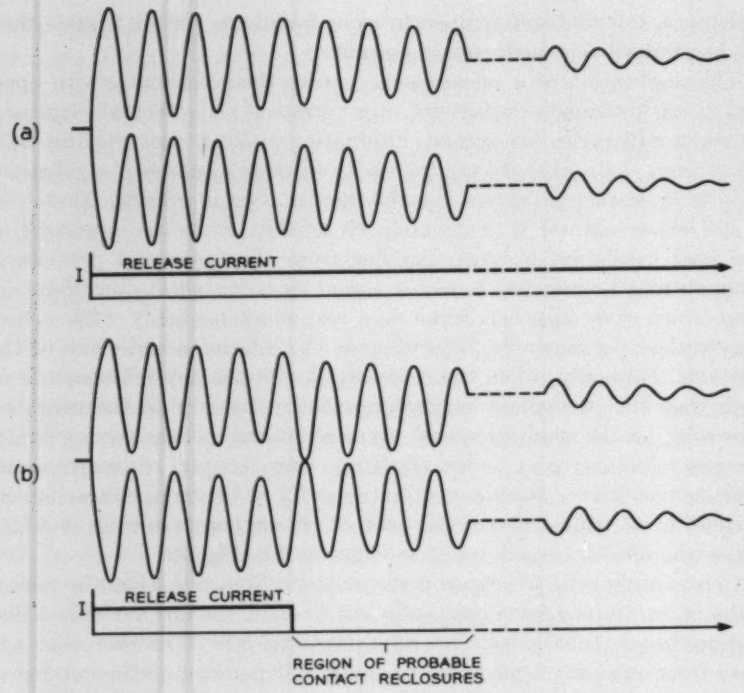


Fig. 12 — Reed motion for (a) long and (b) short release pulse.

not only to have low contact resistance, but also to withstand environmental vibration, to avoid any modulation of current flowing through the contacts that might produce excessive noise in a talking circuit. The contact force of a sealed contact is a function of the reed magnetization and of the inherent sealed contact sensitivity. For a given applied magnetomotive force, a sealed contact that has inherently a high NI-hold value will have a lower contact force than one with a low NI-hold value, because of differences in the force/displacement ratios between the two sealed contact units. Fig. 13 illustrates the relative contact forces for maximum, average and minimum hold value sealed contacts at various levels of hold NI in a test coil containing a single sealed contact.

When a permanent magnet is associated with one or more sealed contacts, the contact forces are no longer a function of the coil NI, because, with zero coil current, the contacts are held closed by the flux

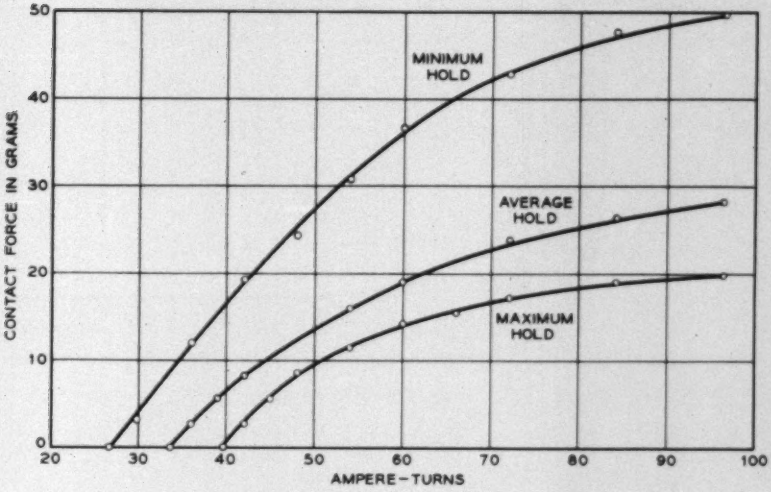


Fig. 13 — Contact force vs. ampere-turns.

supplied by the permanent magnet. Since the permanent magnet strength is adjusted after the relays are completely assembled, it is difficult to know precisely the individual closed reed contact magnetization and the corresponding contact force. However, since contact resistance is a function of contact force, estimates have been derived for the contact force in a magnetic latching relay by graphical analysis of the NI hold versus contact force and the NI hold versus contact resistance parameters of sealed contacts in a neutral-type relay structure and relating these parameters to experimental measurements of contact resistance in magnetic latching relays.

The measurements involved progressively increasing the current in the relay winding in a direction that tends to cause the magnetically latching relay to release. At discrete levels of coil NI, contact resistance measurements were made up to the point where the contacts opened. Then the hold NI for neutral and for magnetically latched relays were related to the common parameter, contact resistance. Fig. 14 illustrates the change in contact resistance of a test universe of sealed contacts tested individually in a test coil, as the coil NI was varied.

From this study it was concluded that under the normal circuit hold condition of zero coil current, the contact force of a sealed contact in a 12 contact magnetic latching relay, for example would be about 13 grams.

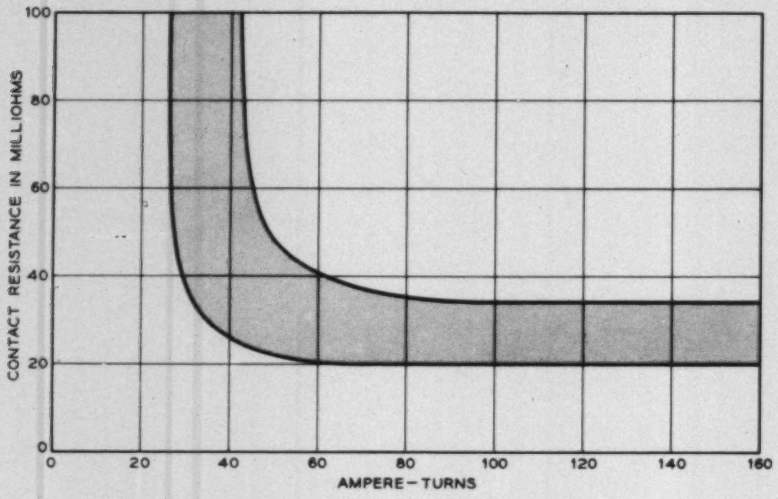


Fig. 14 — Contact resistance vs. ampere-turns.

Contact forces of this order are considered adequate to meet normal central office environmental conditions.

5.3.2 *Vibration*

The zero current hold feature of magnetic latching relays makes them very attractive for applications in equipments to be mounted at locations remote from the central office. In such applications, since the equipments and relays may be subjected to vibration, it is necessary to have assurance of a margin of safety against vibration. Studies were conducted to determine the degree of vibration that was likely to momentarily close open contacts, momentarily open closed contacts and the levels at which it was possible to cause noise at closed contacts in a talking circuit.

Employing a biasing current technique similar to that used for the contact force and contact resistance studies, shake-table tests were made

TABLE I — MARGINS

	4-Contact Relay	12-Contact Relay
Against false contact closures	88%	93%
Against false contact opens	94	93
Against excessive noise	59	56

with four- and 12-contact magnetic latching relays over a range of 0 to 60 cycles at peak-to-peak amplitudes of 0.005 to 0.020 inch. Typical of conservative design for switching application are the margins obtained under the vibration condition of 10 cycles per second at peak-to-peak amplitude of 0.012 inch, as given in Table I. The margins are expressed as a percentage of the nonoperate test and of the hold test values. They indicate that, under the vibration condition imposed on the four-contact relay, for example, a current in a direction tending to operate the relay that was 88 per cent of the specified nonoperate test value was required to induce false contact closures; a current in a direction tending to release the relay that was 94 per cent of the specified test hold value was required to induce false contact opens; a similar current that was 59 per cent of the specified test hold value was required to cause excessive contact noise. The noise measurements were made with the contacts carrying 100 milliamperes DC in the subscriber set talking path of a simulated intraoffice trunk with HA1 receiver weighting. Noise was measured directly across the HA1 receiver with a 2B noise measuring set. Noise levels above 17 dba were considered excessive.

VI. SUMMARY

This paper has reviewed and compared magnet-biased relays generally, and has described specifically the design and performance of a new type of magnetic latching relay that makes use of sealed reed contacts.

VII. ACKNOWLEDGMENTS

The authors wish to acknowledge and thank O. M. Hovgaard for his many helpful comments and suggestions and C. J. Engelhardt, W. Bachle, S. F. Sampson and E. F. Holmes, Jr. for their assistance in conducting the static and dynamic studies of relay characteristics.

REFERENCES

1. Brown, J. T. L. and Pollard, C. E., Balanced Polar Mercury Contact Relay, B.S.T.J., **32**, 1953, p. 1271.
2. Hovgaard, O. M. and Perreault, G. E., Development of Reed Switches and Relays, B.S.T.J., **34**, 1955, p. 309.
3. Peek, R. L. and Wagar, H. N., *Switching Relay Design*, D. Van Nostrand Co., New York, 1955.

Potential Distribution and Capacitance of a Graded p-n Junction

By S. P. MORGAN and F. M. SMITS

(Manuscript received June 6, 1960)

For a graded p-n junction under sufficient forward bias, the usual space-charge approximation to the potential breaks down, and a numerical solution of the differential equation satisfied by the potential is required. A procedure is described which avoids the difficulties associated with direct numerical integration of the stiff differential equation, and which yields a pair of very close upper and lower bounds to the potential at all points. For a linearly graded junction, tables of bounds are given which nowhere differ by as much as 1 per cent, and which effectively bridge the gap between the space-charge case and the neutral case.

The computer solutions are used to calculate the voltage dependence of the stored charge (low-frequency ac capacitance) of a graded junction numerically, as a function of the bias voltage across the junction. The expression for the capacitance is split into two parts, one of which dominates in the neutral case and the other in the space-charge case. With properly normalized variables, it is possible to give a universal plot of small-signal ac capacitance against applied voltage. The results differ from the usual approximate formulas by amounts ranging up to nearly 10 per cent.

I. INTRODUCTION

In his theory of p-n junctions in semiconductors, Shockley[1] shows that the potential distribution in a linearly graded p-n junction satisfies a one-parameter differential equation of the form

$$\frac{d^2U}{dz^2} = \sinh U - Kz. \tag{1}$$

The solution of this nonlinear equation cannot be obtained by analytical methods. However, Shockley gives two approximations, each applicable in an extreme case. The space-charge approximation ($K \gg 1$) is valid in the case of a steep gradient at the junction, while the neutral approxi-

mation ($K \ll 1$) holds for a gentle gradient. Since in almost all cases of practical interest the gradients in p-n junctions are of such a magnitude that the space-charge approximation is valid, it has become the accepted form for the analysis of p-n junctions.

Shockley derived the differential equation (1) for the case of equilibrium — that is, for the case of zero bias across the junction. Recently Moll[2] has shown that the potential distribution in a p-n junction under bias satisfies the same one-parameter differential equation. The parameter K is now a function of the applied voltage, and it turns out that for most practical gradients in p-n junctions under forward bias the space-charge approximation ceases to be valid. For a proper description of the potential distribution in a biased junction, a numerical solution of (1) is required. Such a solution is obtained in the present paper and applied to the computation of the small-signal ac capacitance of the junction as a function of the bias voltage.

First Moll's derivation of the generalized differential equation for the potential distribution in a graded p-n junction under bias is reviewed. There follows a discussion of the physical principles which underlie the approximate solutions and the exact solution. Calculation of the exact solution by numerical integration, however, encounters the difficulty that one boundary condition is specified at infinity, and for large z the solutions found by ordinary numerical integration procedures diverge rapidly from the desired solution. This instability of the desired solution is called *stiffness*. An effective way to calculate the potential in spite of the stiffness consists in the computation of close upper and lower bounds, that is, two very nearby curves between which the actual potential must lie for all z. Mathematical details and extensive numerical tables are contained in Appendix A, while a graphical comparison between the exact space-charge distribution and the distribution assumed in the space-charge approximation is given in the body of the paper.

Finally the computer solutions are used to calculate the voltage dependence of the stored charge (low-frequency ac capacitance) of a graded junction. The results of the computation, which is described in Appendix B, yield a universal plot of the small-signal ac capacitance of the junction against applied voltage.

II. THE GENERALIZED DIFFERENTIAL EQUATION

To obtain the differential equation for the potential distribution in a biased junction, it is convenient to use the model of a p-n junction which was employed by Shockley in his derivation of the space-charge capacitance. The model (Fig. 1) assumes a linearly graded p-n junction in

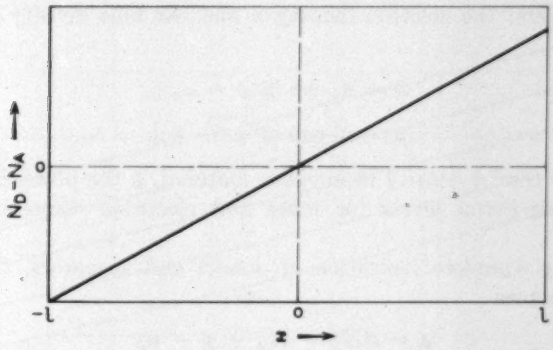


Fig. 1 — The pseudo-equilibrium model of a linearly graded junction.

which the net impurity concentration is given as $N_D - N_A = ax$, with N_D and N_A the density of donors and acceptors, respectively. Contacts are applied at $x = +l$ and $x = -l$. In the model the following processes are imagined to be prevented: (1) electron and hole recombination; (2) electron flow across the p-region contact at $x = -l$; (3) hole flow across the n-region contact at $x = +l$. Under such a pseudo-equilibrium condition, holes which flow in at $x = -l$ must remain in the structure; similarly, electrons flowing in at $x = +l$ must remain inside. No direct current is therefore possible, and the structure behaves like a capacitor.

If the potential of the p-region contact at $x = -l$ is increased by a voltage V with respect to the n-region contact, holes will flow into the specimen until the quasi-Fermi level φ_p for holes has increased by V, so that the holes inside are in equilibrium with the contact supplying the potential. A similar electron flow will occur at $x = +l$. In the pseudo-equilibrium state the quasi-Fermi levels for holes and electrons will have no gradient and will be separated by an amount equal to the applied voltage V.

In an actual p-n junction the region between $x = -l$ and $x = +l$ can be identified with the transition region. Under forward bias the current flow across the transition region gives rise to very small gradients in the quasi-Fermi levels, which can be neglected for a consideration of the potential distribution. Even for the derivation of the voltage-current characteristic, constant quasi-Fermi levels across the transition region are commonly assumed. The use of the pseudo-equilibrium model emphasizes the aspects of the transition region which are being considered here; this is particularly true in the derivation of the capacitance.

To derive the generalized differential equation, Boltzmann statistics

is assumed and the electron density n and the hole density p are expressed by

$$n = n_i \exp [\beta(\psi - \varphi_n)], \tag{2a}$$

$$p = n_i \exp [\beta(\varphi_p - \psi)], \tag{2b}$$

with n_i the carrier density in intrinsic material, ψ the potential, φ_p and φ_n the quasi-Fermi levels for holes and electrons respectively, and $\beta = q/kT$.

Assuming complete ionization of donors and acceptors, the charge density becomes

$$\rho = q[N_D - N_A + p - n]. \tag{3}$$

In view of (2), this can be written as

$$\rho = q\left(N_D - N_A - 2n_i \exp [\tfrac{1}{2}\beta(\varphi_p - \varphi_n)] \sinh \{\beta[\psi - \tfrac{1}{2}(\varphi_p + \varphi_n)]\}\right). \tag{4}$$

If one introduces a normalized potential

$$U = \beta[\psi - \tfrac{1}{2}(\varphi_n + \varphi_p)], \tag{5}$$

and remembers that

$$\varphi_p - \varphi_n = V, \tag{6}$$

one obtains

$$\rho = q(N_D - N_A - 2n_i e^{\beta V/2} \sinh U). \tag{7}$$

To find the potential distribution, this charge density has to be inserted into Poisson's equation

$$\frac{d^2\psi}{dx^2} = -\frac{\rho}{\epsilon}. \tag{8}$$

On making the substitutions

$$\frac{1}{\mathfrak{L}_D^2} = \frac{2n_i\beta q}{\epsilon}, \tag{9}$$

$$z = \frac{x}{\mathfrak{L}_D} e^{\beta V/4}, \tag{10}$$

$$N_D - N_A = ax, \tag{11}$$

$$\frac{a\mathfrak{L}_D}{2n_i} = K_0, \tag{12}$$

$$K = K_0 e^{-3\beta V/4}, \tag{13}$$

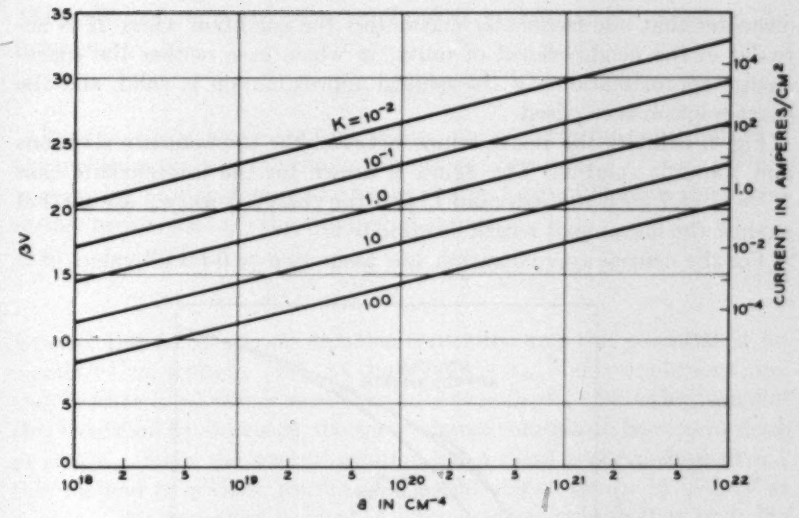


Fig. 2 — The parameter K as a function of forward bias and gradient at the junction.

one obtains for Poisson's equation the one-parameter differential equation

$$\frac{d^2U}{dz^2} = \sinh U - Kz. \tag{14}$$

For the case of zero bias, this differential equation corresponds to the one given by Shockley.* He gave two limiting approximations to the solution, namely the space-charge approximation for $K \gg 1$ and the neutral approximation for $K \ll 1$. Note, however, that on account of the factor $e^{-3\beta V/4}$ in the general expression for K as given by (13), K can pass through unity with increasing forward bias even if $K_0 \gg 1$.

To illustrate the effect of forward bias on K, Fig. 2, which is reproduced from Moll,[2] shows the values of K for practical gradients at the junction (values of a between 10^{18} and 10^{22} cm^{-4}) and for various values of forward bias. Also shown are reasonable estimates of current densities corresponding to such forward biases in practical cases. This plot dem-

* Shockley uses the normalized length coordinate $y = Kz$, so that (14) takes the form

$$\frac{d^2U}{dy^2} = \frac{1}{K^2}(\sinh U - y).$$

onstrates that one frequently encounters the condition where K is actually in the neighborhood of unity, in which case neither the space-charge approximation nor the neutral approximation is valid, and the exact solution is required.

Fig. 3 exhibits the relationships between the approximate solutions and the true solution. The figure is drawn for the intermediate case $K = 5$, and both the potential U and the charge density ρ are plotted against the normalized length coordinate z.

For the neutral approximation, one assumes $\rho \approx 0$ for all values of z.

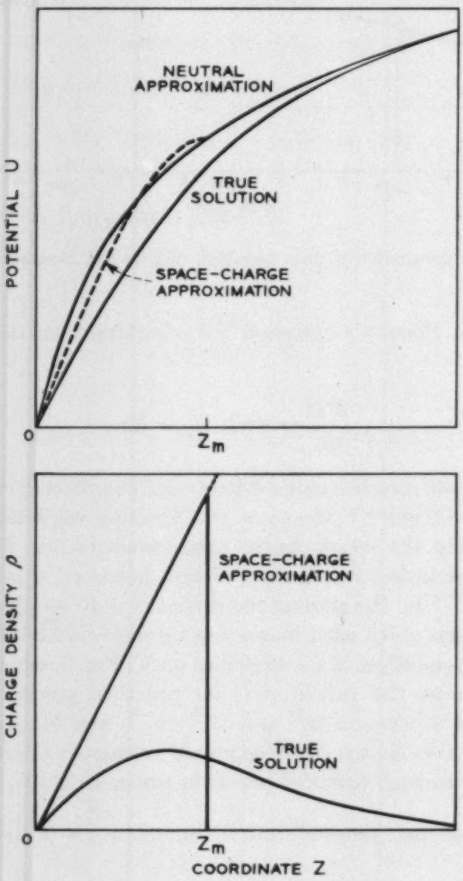


Fig. 3 — Relationships between the approximate solutions and the true solution for the case $K = 5$.

Then $d^2U/dz^2 \approx 0$, and the potential is approximately equal to the function

$$\bar{U} = \sinh^{-1}Kz. \tag{15}$$

For the space-charge approximation, one neglects the contribution of the mobile carriers to the charge density out to a point $|z| = z_m$, which is regarded as the edge of the space-charge region. Thus in the space-charge region one uses the differential equation

$$\frac{d^2U}{dz^2} = -Kz, \qquad |z| < z_m. \tag{16}$$

Beyond the point z_m one assumes neutrality and the potential is set equal to $\bar{U}$ as given by (15). At the point $z = z_m$, the two solutions join. One further condition is necessary to determine z_m; Shockley supplies this condition by requiring the space-charge solution to have zero slope at $z = z_m$. Since the neutral solution has a small positive slope at z_m, this method of joining results in a slight discontinuity in dU/dz at $z = z_m$. An improved method of joining would require that both the function and the first derivative be continuous at $z = z_m$. However, such a condition complicates the algebra without significantly improving the accuracy of the approximation.

The true solution is asymptotic to the neutral solution $\bar{U}$ as $z \to \infty$. The space-charge distribution will not have a sharp boundary, as is assumed in the space-charge approximation. A fraction of the fixed charges associated with the impurities are compensated by mobile carriers, giving rise to an actual net charge density like that indicated in the lower part of Fig. 3. Since for the exact solution the charge is distributed over a larger distance, the total charge must be smaller than is assumed in the space-charge approximation.

Calculation of the exact potential distribution in the junction involves numerical solution of the differential equation (14) subject to the boundary conditions $U = 0$ at $z = 0$ and $U \to \sinh^{-1}Kz$ as $z \to \infty$. The method of computation is described in Appendix A, where tables of upper and lower bounds for U which nowhere differ by as much as 1 per cent are given for 16 values of K ranging from 0.1 to 10,000.

The results of the calculations, for selected values of K, are also shown in terms of charge densities in Fig. 4. In this figure the fractional compensation of the fixed charges by mobile carriers, namely the ratio $(n - p)/(N_D - N_A)$, is plotted against z/z_m, where z_m is the space-charge width for the space-charge approximation. The value of z_m is plotted against K in Fig. 5. The normalization by z_m allows a complete representation of the results in one figure. It also readily permits a com-

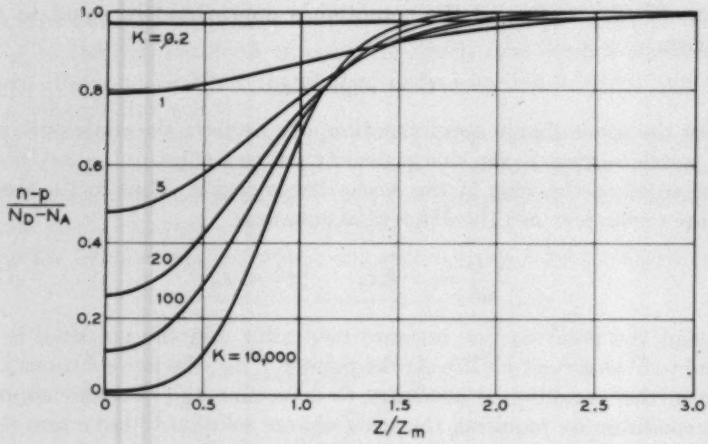


Fig. 4 — Fractional compensation of fixed charges by mobile carriers for various values of K.

parison between the true solution and the space-charge approximation, since for the latter the ordinate is zero for $z < z_m$ and unity for $z > z_m$.

It is apparent from Fig. 4 that the true solution differs markedly from the space-charge approximation, even for values of K as large as 10^4.

III. THE CAPACITANCE OF A GRADED p-n JUNCTION

The significant difference between the charge distribution assumed in the space-charge approximation and the actual charge distribution

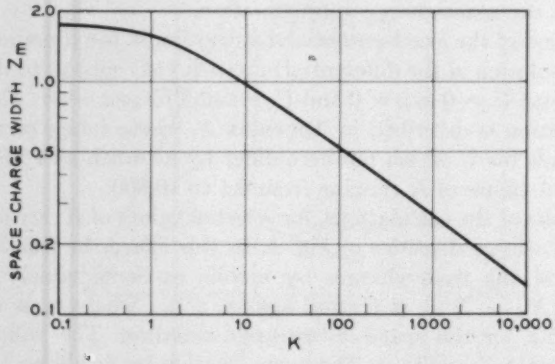


Fig. 5 — The space-charge width for the space-charge approximation.

prompted us to compute the voltage dependence of the stored charge (low-frequency ac capacitance) of a graded junction. For this problem the pseudo-equilibrium model is particularly helpful, since only capacitive currents will flow. A change in the voltage between the two contacts at $x = -l$ and $x = +l$ is associated with a change in the total number of mobile carriers stored in the structure. The total number of electrons and the total number of holes are always equal. The two types of carriers are equivalent to the charge on the two "plates" of the capacitor, and by considering the total number of holes one can express the capacitance as

$$C = \frac{d}{dV} q \left(\int_{-l}^{+l} p \, dx \right). \tag{17}$$

Because of the symmetry of the junction, the expression (17) can be written

$$C = \frac{d}{dV} \left(q \int_0^l (p + n) \, dx \right). \tag{18}$$

Since the density of fixed charges associated with the impurities is voltage-independent, one can rewrite this equation as

$$C = \frac{d}{dV} \left(q \int_0^l [(p + n) - (N_D - N_A)] \, dx \right). \tag{19}$$

After eliminating $N_D - N_A$ by (3), one obtains as the expression for the capacitance

$$C = \frac{d}{dV} \left(2q \int_0^l p \, dx \right) - \frac{d}{dV} \left(\int_0^l \rho \, dx \right). \tag{20}$$

The capacitance accordingly consists of two parts: the "neutral" capacitance

$$C_n = \frac{d}{dV} \left(2q \int_0^l p \, dx \right), \tag{21}$$

and the "space-charge" capacitance

$$C_\rho = -\frac{d}{dV} \left(\int_0^l \rho \, dx \right). \tag{22}$$

Since according to (13) the parameter K depends on V through the factor $e^{-3\beta V/4}$, one may express derivatives with respect to V in terms of derivatives with respect to K. Furthermore the expression for the space-charge capacitance converges very strongly as $l \to \infty$. Thus only a negligible error will be made if one replaces the upper limit of the in-

tegral by infinity. For the neutral capacitance, however, it is necessary to retain a finite value of l. After expressing p and ρ in terms of the potential by (2b) and (8) respectively, and transforming to dimensionless variables, one gets

$$C_n = \frac{C_0}{K^{\frac{1}{2}}}\left[\int_0^{z_l} e^{-U}dz + z_l e^{-U_l} - 3\frac{\partial}{\partial(\log K)}\int_0^{z_l} e^{-U}dz\right], \quad (23)$$

$$C_\rho = \frac{C_0}{K^{\frac{1}{2}}}\frac{\partial}{\partial z}\left[3\frac{\partial U}{\partial(\log K)} - U\right]_{z=0}. \quad (24)$$

In both cases C_0 is a normalizing capacitance given by

$$C_0 = \frac{\epsilon K_0^{\frac{1}{2}}}{4\mathfrak{L}_D}, \quad (25)$$

which involves only material constants and the gradient at the junction. For the space-charge capacitance, the expression multiplying C_0 is a function of K only. For the neutral capacitance, the multiplying function contains in addition the upper limit of the integral in (21), which can be expressed either by the actual coordinate l, by the normalized coordinate z_l, or by the normalized potential U_l of the electrode.

For $K \to 0$ the space-charge capacitance disappears and the relationship between U and K is given by (15). The neutral capacitance in this limit corresponds to the "capacitance for the neutral case" as described by Shockley,* and it takes the form

$$C_n \underset{K\to 0}{\approx} \frac{2U_l}{K^{\frac{1}{2}}} C_0. \quad (26)$$

For $K \gg 1$ the space-charge capacitance is much greater than the neutral capacitance. By using in (24) the functional relationship between U and K that results from the space-charge approximation, one obtains for the space-charge capacitance

$$C_\rho \approx \frac{2(3U_m)^{\frac{1}{2}}C_0}{3U_m \coth U_m - 1} \underset{K\to\infty}{\approx} \frac{2C_0}{(3U_m)^{\frac{1}{2}}}, \quad (27)$$

* In the present notation, Eq. (2.41) of Ref. 1 is equivalent to

$$C = \frac{(2U_l - 1)C_0}{K^{\frac{1}{2}}}.$$

A consistent retention of terms of order 1 in Shockley's derivation would, however, have led to a result exactly equivalent to (26) above. In particular it is necessary to write, for large x,

$$\exp(2\sinh^{-1} x) \approx 4x^2 + 2,$$

whereas Shockley effectively approximates this function by $4x^2$.

with U_m the normalized potential at the edge of the space-charge region, which is related to K by

$$K = \left(\frac{\sinh^3 U_m}{3U_m}\right)^{\frac{1}{2}}. \tag{28}$$

If one approximates $U_m = U_{m0} - \beta V/2$, where U_{m0} is the normalized potential at the edge of the space-charge region for zero bias, the last term of (27) agrees with the space-charge capacitance given by Shockley in Eq. (2.45) of Ref. 1.

The functions multiplying C_0 in (23) and (24) have been evaluated numerically, as described in Appendix B. In the expression for C_n a value of 10 has been used for U_l. (It should be emphasized that a fixed U_l does not correspond to a fixed l!) The results are shown in Fig. 6. Also shown are the approximations (26) and (27) for C_n and C_ρ respectively.

It may be noted that for large K the values for the exact space-charge capacitance fall above the approximate values. This at first appears surprising, since as mentioned above the total charge in the transition region is actually less than assumed under the space-charge approximation. However if one remembers that the validity of the approximation increases with increasing K, it becomes plausible that the change in total charge with K must be larger for the exact solution.

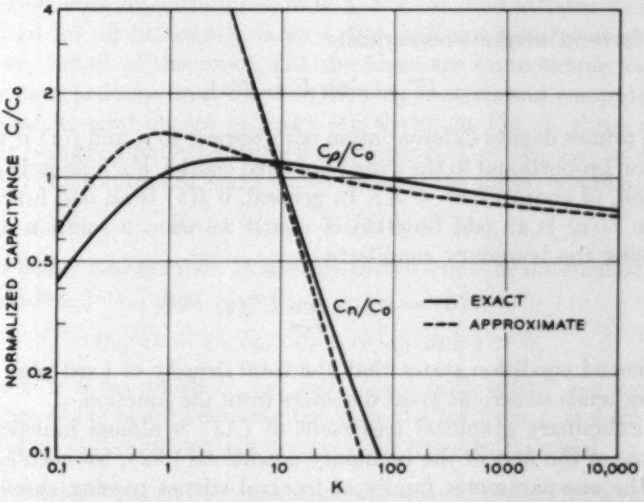


Fig. 6 — Approximate and exact values of the normalized neutral capacitance and the normalized space-charge capacitance.

The abscissa in Fig. 6 can be considered as a linear scale in $-V$, the applied voltage, with zero coinciding with the point $K = K_0$. Thus Fig. 6 may be used as a plot of the small-signal ac capacitance against voltage. In particular, it is apparent from Fig. 6 that in the range in which the neutral capacitance dominates, the approximation given by (26) is a good description of this capacitance. This finding is important, since in practical cases the value of U_l needs to be specified. One would choose as the boundary of the transition region a point up to which the assumption of constant quasi-Fermi levels is reasonably correct. Beyond this point one has to use the continuity equation to find the "diffusion" capacitance, which must be added to the capacitance of the transition region.

IV. ACKNOWLEDGMENTS

We wish to thank R. C. Prim for stimulating discussions of the mathematical aspects of this problem and for the use of his unpublished results. The numerical work has also had the benefit of unpublished investigations by A. Kooharian and W. L. Miranker.

APPENDIX A

Calculation of the Potential Distribution

We have to integrate numerically

$$U'' = \sinh U - f(z), \tag{A1}$$

where primes denote differentiation with respect to z, and $f(z)$ is a given function proportional to the density of fixed charge. For a linearly graded junction, of course, $f(z) = Kz$. In general, if $f(z)$ is an odd function of z, then $U(z)$ is an odd function of z, and we want a solution of (A1) satisfying the boundary conditions

$$U(0) = 0, \qquad \lim_{z \to \infty} U''(z) = 0. \tag{A2}$$

The second condition states that the total density of fixed plus mobile charges tends to zero at great distances from the junction.

A preliminary graphical treatment of (A1) is almost indispensable. Because of the first of the boundary conditions (A2), we need consider only the one-parameter family of integral curves passing through the origin with arbitrary slope. In view of the second boundary condition,

the problem is to find an integral curve of (A1) which passes through the origin with such a slope that at infinity it approaches the curve

$$\tilde{U}(z) = \sinh^{-1} f(z); \tag{A3}$$

and in fact $U(z)$ must approach $\tilde{U}(z)$ fast enough so that

$$\lim_{z \to \infty} [\sinh U(z) - \sinh \tilde{U}(z)] = 0. \tag{A4}$$

One might, perhaps, be willing to take the existence of a physically meaningful solution of the boundary value problem for granted, on the ground that no matter how the fixed charge density $f(z)$ varies with position, the holes and electrons should be able to distribute themselves so that equilibrium is produced. A formal proof that the problem does indeed have a unique solution can be given under the assumptions that $f(z)$ is a nonnegative function with two continuous derivatives for $0 \leqq z < \infty$, that $\tilde{U}(z)$ is everywhere concave downward, i.e.,

$$\tilde{U}''(z) \leqq 0 \qquad \text{for} \quad 0 \leqq z < \infty, \tag{A5}$$

and that

$$\lim_{z \to \infty} \tilde{U}''(z) = 0. \tag{A6}$$

Doubtless it would be sufficient to assume that $f(z)$ has only a finite number of finite discontinuities in $0 \leqq z < \infty$, and satisfies the given conditions for all sufficiently large z. We shall not take space to write out every detail of the proof, but the ideas are quite simple and will now be sketched as a basis for the following analysis and computations.

Typical integral curves of (A1) are shown in Fig. 7, these curves actually being computer solutions for the case $f(z) = 10z$. Let us denote by $U_s(z)$ the solution of (A1) which satisfies the initial conditions

$$U_s(0) = 0, \qquad U_s'(0) = s, \tag{A7}$$

where s is any real number. It is easily shown from the differential equation that if $s_2 > s_1$, then

$$U_{s_2}(z) - U_{s_1}(z) > (s_2 - s_1) \sinh z > 0, \tag{A8}$$

for any $z > 0$ for which $U_{s_2}(z)$ and $U_{s_1}(z)$ both exist.* We see from (A8) that $U_s(z)$ is an increasing function of s for fixed z, and that any two integral curves ultimately diverge with exponential speed.

* The solution $U_s(z)$ may become infinite as z approaches some finite value z_s, and so may not exist for $z \geqq z_s$. We shall not be concerned with such movable singularities, except to recognize that they can occur.

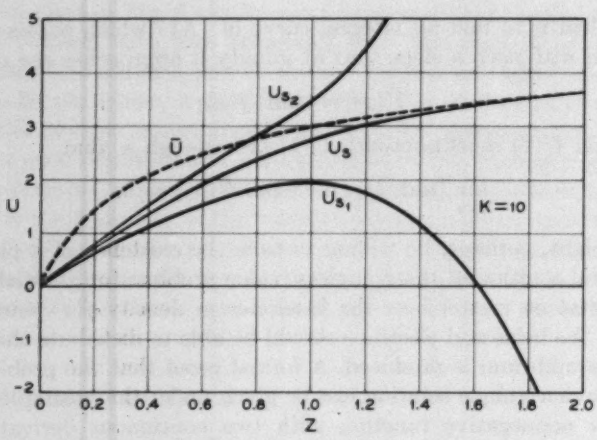


Fig. 7 — Integral curves of the differential equation $U'' = \sinh U - Kz$.

An integral curve which once crosses $\bar{U}(z)$ can never recross, since it has positive second derivative thenceforth, while by assumption $\bar{U}(z)$ has negative second derivative. Similarly, an integral curve which crosses the z-axis going downward has negative second derivative thenceforth, and can never recross the axis.

Now let us imagine, as is always possible, that we start with an initial slope s large enough so that $U_s(z)$ does cross $\bar{U}(z)$, and that we gradually decrease s until we come to the first integral curve that does not cross $\bar{U}(z)$ for any z. On the other hand we could start with a negative value of s, so that $U_s(z)$ surely lies below the z-axis, and increase s until we reach the first integral curve that does not lie below the z-axis anywhere. It turns out that these two distinguished integral curves are actually one and the same, so that the differential equation has exactly one solution which lies between $U = \bar{U}(z)$ and $U = 0$ for all z. (If there were two such solutions, by (A8) the distance between them would ultimately exceed $\bar{U}$.) We shall call this solution $U_S(z)$. It may be proved that $U_S(z)$ approaches $\bar{U}(z)$ and satisfies the boundary condition (A4) at infinity.

The object from now on is to find upper and lower bounds for the desired solution; that is, neighboring functions $U_+(z)$ and $U_-(z)$ such that

$$U_-(z) \leqq U_S(z) \leqq U_+(z) \qquad \text{for} \quad z \geqq 0. \tag{A9}$$

Clearly, if $s_1 < S < s_2$ then $U_{s_1}(z)$ is a lower bound and $U_{s_2}(z)$ is an upper bound; but according to (A8) these two curves are infinitely

far apart for infinite z. From the preceding discussion we know that $U_+(z) = \tilde{U}(z)$ is always an upper bound and $U_-(z) = 0$ is always a lower bound. Furthermore, if $U_{1+}(z)$ and $U_{2+}(z)$ are upper bounds, then $\min [U_{1+}(z), U_{2+}(z)]$ is an upper bound; and if $U_{1-}(z)$ and $U_{2-}(z)$ are lower bounds, then $\max [U_{1-}(z), U_{2-}(z)]$ is a lower bound.

Given a lower bound $U_-(z)$, an upper bound may be constructed by two quadratures, as follows: Let

$$\begin{aligned} \eta''(z) &= \sinh U_-(z) - f(z), \\ \eta(0) &= 0, \\ \eta'(0) &= s, \end{aligned} \tag{A10}$$

and choose the initial slope s so that for some $z_m > 0$ we have

$$\eta(z_m) = \tilde{U}(z_m). \tag{A11}$$

The upper bound is then

$$U_+(z) = \begin{cases} \eta(z), & 0 \leqq z \leqq z_m, \\ \tilde{U}(z), & z > z_m. \end{cases} \tag{A12}$$

To prove that $\eta(z) \geqq U_s(z)$ for $0 \leqq z \leqq z_m$, we observe that the inequality is certainly satisfied at $z = 0$ and at $z = z_m$. If there were a subinterval in which $\eta(z) < U_s(z)$, then at some point z_1 of the subinterval we should have $\eta''(z_1) > U_s''(z_1)$, which is impossible in view of the differential equations (A1) and (A10). The contradiction establishes that $\eta(z)$ is an upper bound in $0 \leqq z \leqq z_m$.

Similarly, from a given upper bound $U_+(z)$ we can construct a lower bound by taking

$$\begin{aligned} \eta''(z) &= \sinh U_+(z) - f(z), \\ \eta(0) &= 0, \\ \eta'(0) &= s, \end{aligned} \tag{A13}$$

and choosing s so that for some $z_0 > 0$ we have

$$\eta'(z_0) = 0. \tag{A14}$$

The lower bound is then

$$U_-(z) = \begin{cases} \eta(z), & 0 \leqq z \leqq z_0, \\ \eta(z_0), & z > z_0; \end{cases} \tag{A15}$$

but it is not a very good bound because it is ultimately constant, while $U_s(z)$ may tend to infinity for large z.

The Shockley space-charge approximation may be obtained as an elementary upper bound, using (A10). Setting $U_-(z) = 0$, we get for a linearly graded junction with $f(z) = Kz$ an upper bound of the form

$$U_+(z) = \begin{cases} sz - Kz^3/6, & 0 \leqq z \leqq z_m, \\ \sinh^{-1} Kz, & z > z_m, \end{cases} \tag{A16}$$

where

$$sz_m - Kz_m^3/6 = \sinh^{-1} Kz_m. \tag{A17}$$

Since both s and z_m are disposable parameters, a second relationship may be imposed between them. Shockley requires the slope of the upper bound to vanish* at $z_m - 0$, so that

$$s - Kz_m^2/2 = 0. \tag{A18}$$

The solutions of (A17) and (A18) are then expressed, in terms of a parameter U_m which satisfies

$$\left(\frac{\sinh^3 U_m}{3U_m}\right)^{\frac{1}{2}} = K, \tag{A19}$$

by the equations

$$\begin{aligned} z_m &= \frac{\sinh U_m}{K}, \\ s &= \frac{\sinh^2 U_m}{2K}. \end{aligned} \tag{A20}$$

The space-charge width z_m, which was plotted in Fig. 5, is tabulated in Table I as a function of K, together with the initial slope s according to the space-charge approximation, and the initial slope S obtained by the numerical method described below.

Unfortunately it does not appear to be possible to converge on the desired solution $U_S(z)$ by repeated applications of (A10) and (A13). A limited amount of numerical experimentation suggests that iteration, even starting from a bound which is known to be very close to $U_S(z)$, rapidly leads to a particular pair of curves which are a finite distance apart for every $z > 0$. Although it is possible, as Prim[3] and Morrison[4] have shown, to obtain a number of analytic upper and lower bounds for $U_S(z)$ of varying degrees of complexity, the most practical way of actually computing close upper and lower bounds seems to be in terms of integral curves of the differential equation itself. Thus let $U_{s_2}(z)$ be an

* A slightly better bound would be obtained by making the slope of $U_+(z)$ continuous at z_m.

TABLE I—SPACE-CHARGE WIDTH AND INITIAL SLOPE OF $U(z)$

K	Space-Charge Approximation		Exact Solution
	z_m	s	S
0.1	1.7278	0.14926	0.099094
0.2	1.7158	0.29440	0.19402
0.5	1.6535	0.68348	0.44698
1	1.5396	1.1852	0.78927
2	1.3760	1.8935	1.3134
5	1.1350	3.2206	2.3792
10	0.9610	4.6180	3.5599
20	0.8046	6.4737	5.1709
50	0.6287	9.8810	8.1895
100	0.5183	13.432	11.379
200	0.4255	18.108	15.614
500	0.3262	26.610	23.378
1000	0.2660	35.389	31.449
2000	0.2165	46.864	42.049
5000	0.1644	67.568	61.264
10000	0.1333	88.791	81.052

integral curve of (A1) which passes through the origin with some initial slope s_2 and crosses $\bar{U}(z)$ at some point z_0. Then an upper bound for $U_S(z)$ is given by

$$U_+(z) = \begin{cases} U_{s_2}(z), & 0 \leqq z \leqq z_0, \\ \bar{U}(z), & z > z_0. \end{cases} \tag{A21}$$

Similarly a lower bound may be constructed from any integral curve $U_{s_1}(z)$ which starts with a positive slope s_1 but ultimately turns downward. Ordinarily we do not have to follow $U_{s_1}(z)$ beyond the point where its slope is zero. Let us introduce the auxiliary function

$$U_{-1}(z) = \sinh^{-1}[f(z) + \bar{U}''(z)]. \tag{A22}$$

$U_{-1}(z)$ is a known function of z which approaches $\bar{U}(z)$ from below for large z as $\bar{U}''(z)$ approaches zero; it is plotted in Fig. 8 for the case $f(z) = 10z$. Now suppose that for some positive z_0 we have

$$\bar{U}(z) - U_{-1}(z) \leqq \epsilon_0 \quad \text{for} \quad z \geqq z_0, \tag{A23}$$

where

$$\epsilon_0 = \bar{U}(z_0) - U_{s_1}(z_0). \tag{A24}$$

Then a lower bound for $U_S(z)$ is given by

$$U_-(z) = \begin{cases} U_{s_1}(z), & 0 \leqq z \leqq z_0, \\ \bar{U}(z) - \epsilon_0, & z > z_0. \end{cases} \tag{A25}$$

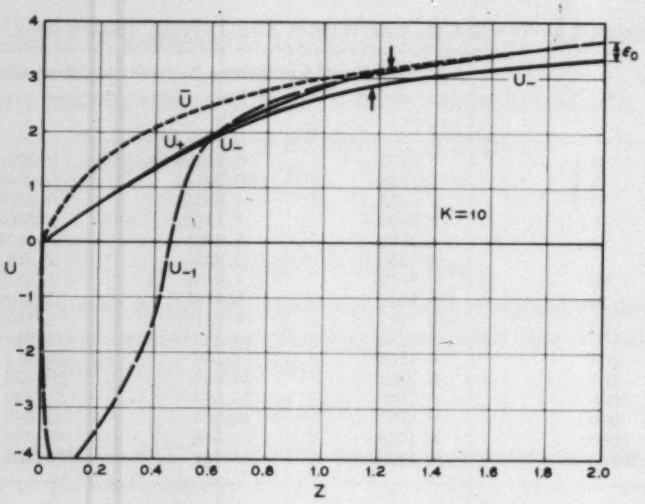


Fig. 8 — Upper and lower bounds, together with the curves $\bar{U}(z)$ and $U_{-1}(z)$.

In other words, $U_s(z)$ lies in a strip of constant width ϵ_0 below $\bar{U}(z)$ for $z \geqq z_0$.

To prove the last statement, define

$$\epsilon(z) = \bar{U}(z) - U_s(z). \tag{A26}$$

Since $U_{s_1}(z)$ ultimately turns downward, we know that it lies below $U_s(z)$ at $z = z_0$. We also know that $U_s(z)$ approaches $\bar{U}(z)$ at infinity, so

$$\epsilon(z_0) < \epsilon_0 \qquad \text{and} \qquad \lim_{z \to \infty} \epsilon(z) = 0. \tag{A27}$$

It follows that if there were some interval beyond z_0 in which $\epsilon(z) > \epsilon_0$, there would be some point z_1 of the interval at which $\epsilon''(z_1) < 0$. But at this point we have

$$\begin{aligned} \epsilon''(z_1) &= \bar{U}''(z_1) - U_s''(z_1) \\ &= \sinh U_{-1}(z_1) - \sinh U_s(z_1) > 0, \end{aligned} \tag{A28}$$

since, by (A23), $U_s(z_1) < U_{-1}(z_1)$ if $\bar{U}(z_1) - U_s(z_1) > \epsilon_0$. This contradiction proves the theorem.

In practice it is usually possible to choose z_0 to minimize the distance between $U_{s_1}(z)$ and $\bar{U}(z)$, i.e., so that

$$U'_{s_1}(z_0) = \bar{U}'(z_0). \tag{A29}$$

However, this condition is not required for the preceding proof.

Representative upper and lower bounds obtained in this way for $K = 10$ are shown in Fig. 8, together with the curves $\bar{U}(z)$ and $U_{-1}(z)$. For ease in plotting, the value of ϵ_0 has been chosen larger than would be necessary to accept in a practical case. The vertical arrows indicate, respectively, the point at which U_+ becomes $\bar{U}$ and the point at which U_- becomes $\bar{U} - \epsilon_0$.

To mechanize the computation of upper and lower bounds, one has only to integrate (A1) numerically, starting from the origin with a given slope, out to the point where the integral curve identifies itself either by crossing $\bar{U}(z)$ or by attaining zero slope. The initial slope is then adjusted (downward in the former case, upward in the latter), and the process repeated until the bounds nowhere differ by more than a preassigned amount. This approach may be called the "sweep method."

Any standard procedure may be used for the numerical integration. We took the fourth-order Runge-Kutta method,[5] adapted for an equation of the form $y'' = f(x, y)$. In practice the IBM 704 was supplied with the value of K, an initial guess at the slope s, and criteria for stopping; integral curves could be run off at the rate of several per minute.

The criterion for refinement of a particular solution was more or less arbitrarily taken to be the determination of the initial slope to five significant figures.* This led to upper and lower bounds which differ by less than 1 per cent of their mean value in all cases. Typical plots are shown in Figs. 9, 10, and 11. In Fig. 9, where K has the small value 0.1, the solution $U_S(z)$ nowhere differs from $\bar{U}(z)$ by more than 1 per cent, and no attempt has been made to plot the two curves separately. Fig. 10 represents the intermediate value $K = 10$, and Fig. 11 the large value $K = 10{,}000$. The difference between upper and lower bounds for a given K, while obvious in the tabulated values near the upper end of the range of z, is too small to be conveniently plotted on these figures.

Table II shows $\bar{U}(z)$, $U_+(z)$, and $U_-(z)$ for 16 values of K ranging from 0.1 to 10,000. The values of $U_+(z)$ and $U_-(z)$, considered as numerical solutions of the differential equation (A1), are believed to be accurate, with a few possible exceptions, to the four decimal places shown.

* More precisely, $U_+'(0)$ exceeds $U_-'(0)$ by one unit of the fifth significant figure, where $U_+'(0)$ is given as S in Table I.

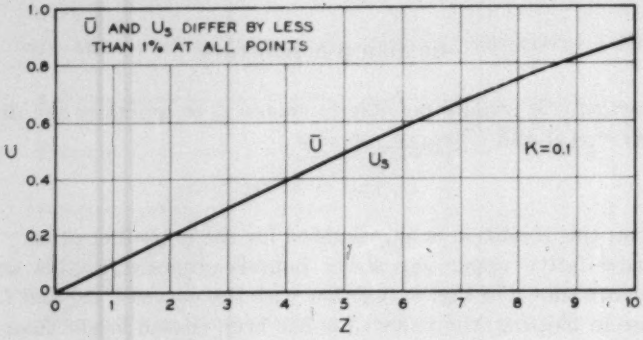


Fig. 9 — Potential distribution $U(z)$ for $K = 0.1$.

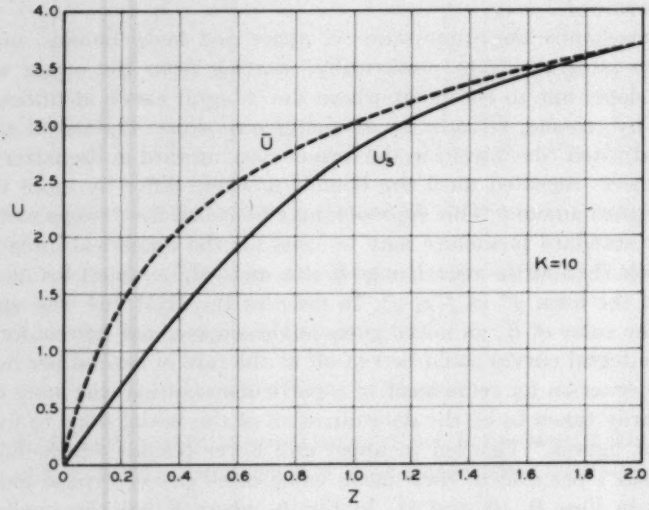


Fig. 10 — Potential distribution $U(z)$ for $K = 10$.

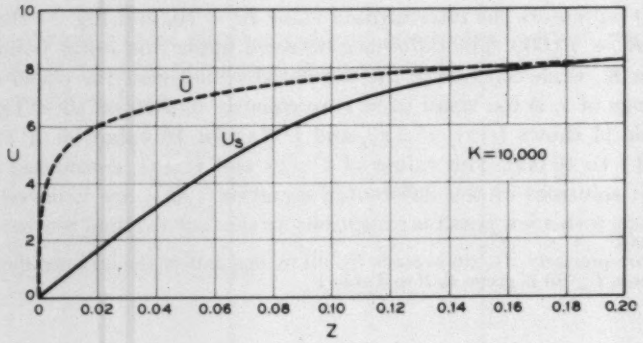


Fig. 11 — Potential distribution $U(z)$ for $K = 10{,}000$.

TABLE II — THE FUNCTIONS $\bar{U}(z)$, $U_+(z)$, AND $U_-(z)$

$K = 0.1$				$K = 0.2$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.	0.	0.	0.	0.
0.500	0.0500	0.0495	0.0495	0.200	0.0400	0.0388	0.0388
1.000	0.0998	0.0989	0.0989	0.400	0.0799	0.0775	0.0775
1.500	0.1494	0.1481	0.1481	0.600	0.1197	0.1162	0.1162
2.000	0.1987	0.1970	0.1970	0.800	0.1593	0.1547	0.1547
2.500	0.2475	0.2454	0.2454	1.000	0.1987	0.1930	0.1930
3.000	0.2957	0.2933	0.2933	1.200	0.2378	0.2311	0.2311
3.500	0.3432	0.3406	0.3406	1.400	0.2765	0.2690	0.2689
4.000	0.3900	0.3872	0.3872	1.600	0.3148	0.3065	0.3064
4.500	0.4360	0.4331	0.4331	1.800	0.3526	0.3436	0.3436
5.000	0.4812	0.4782	0.4782	2.000	0.3900	0.3804	0.3804
5.500	0.5255	0.5224	0.5224	2.200	0.4269	0.4168	0.4167
6.000	0.5688	0.5657	0.5657	2.400	0.4633	0.4527	0.4526
6.500	0.6112	0.6081	0.6081	2.600	0.4990	0.4881	0.4881
7.000	0.6527	0.6496	0.6496	2.800	0.5342	0.5231	0.5230
7.500	0.6931	0.6902	0.6901	3.000	0.5688	0.5576	0.5575
8.000	0.7327	0.7299	0.7297	3.200	0.6028	0.5915	0.5914
8.500	0.7712	0.7686	0.7683	3.400	0.6362	0.6249	0.6248
9.000	0.8089	0.8066	0.8059**	3.600	0.6690	0.6578	0.6576
9.500	0.8456	0.8437	0.8426**	3.800	0.7011	0.6901	0.6898
10.000	0.8814	0.8803	0.8784**	4.000	0.7327	0.7218	0.7215
10.500	0.9163	0.9163*	0.9134**	4.200	0.7636	0.7530	0.7526
				4.400	0.7939	0.7837	0.7832
				4.600	0.8237	0.8138	0.8132
				4.800	0.8528	0.8434	0.8426
				5.000	0.8814	0.8724	0.8715
				5.200	0.9094	0.9010	0.8998
				5.400	0.9368	0.9290	0.9275
				5.600	0.9637	0.9567	0.9547
				5.800	0.9901	0.9838	0.9814
				6.000	1.0160	1.0106	1.0075
				6.200	1.0413	1.0371	1.0331
				6.400	1.0662	1.0633	1.0581
				6.600	1.0906	1.0894	1.0827
				6.800	1.1145	1.1145*	1.1067
				7.000	1.1380	1.1380*	1.1302**
* $\bar{U}(z)$		** $\bar{U}(z) - 0.0029$		* $\bar{U}(z)$		** $\bar{U}(z) - 0.0078$	

$K = 0.5$				$K = 1.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.	0.	0.	0.	0.
0.200	0.0998	0.0893	0.0893	0.200	0.1987	0.1576	0.1576
0.400	0.1987	0.1782	0.1782	0.400	0.3900	0.3135	0.3135
0.600	0.2957	0.2663	0.2663	0.600	0.5688	0.4662	0.4661
0.800	0.3900	0.3532	0.3531	0.800	0.7327	0.6142	0.6142
1.000	0.4812	0.4384	0.4384	1.000	0.8814	0.7564	0.7563
1.200	0.5688	0.5218	0.5218	1.200	1.0160	0.8918	0.8918
1.400	0.6527	0.6030	0.6030	1.400	1.1380	1.0198	1.0198
1.600	0.7327	0.6819	0.6819	1.600	1.2490	1.1401	1.1401
1.800	0.8089	0.7582	0.7581	1.800	1.3504	1.2526	1.2526

TABLE II — *Continued*

$K = 0.5$				$K = 1.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
2.000	0.8814	0.8318	0.8317	2.000	1.4436	1.3574	1.3573
2.200	0.9503	0.9026	0.9026	2.200	1.5297	1.4547	1.4546
2.400	1.0160	0.9707	0.9706	2.400	1.6094	1.5450	1.5450
2.600	1.0785	1.0360	1.0359	2.600	1.6837	1.6289	1.6288
2.800	1.1380	1.0985	1.0984	2.800	1.7532	1.7068	1.7066
3.000	1.1948	1.1584	1.1583	3.000	1.8184	1.7792	1.7790
3.200	1.2490	1.2157	1.2155	3.200	1.8799	1.8468	1.8465
3.400	1.3008	1.2705	1.2703	3.400	1.9379	1.9101	1.9097
3.600	1.3504	1.3230	1.3227	3.600	1.9928	1.9694	1.9688
3.800	1.3980	1.3732	1.3728	3.800	2.0450	2.0253	2.0245
4.000	1.4436	1.4213	1.4208	4.000	2.0947	2.0780	2.0769
4.200	1.4875	1.4675	1.4668	4.200	2.1421	2.1281	2.1263
4.400	1.5297	1.5118	1.5109	4.400	2.1874	2.1757	2.1731
4.600	1.5703	1.5543	1.5532	4.600	2.2308	2.2213	2.2173
4.800	1.6094	1.5953	1.5937	4.800	2.2724	2.2651	2.2591**
5.000	1.6472	1.6349	1.6327	5.000	2.3124	2.3074	2.2991**
5.200	1.6837	1.6731	1.6700	5.200	2.3509	2.3487	2.3376**
5.400	1.7191	1.7101	1.7059	5.400	2.3880	2.3880*	2.3747**
5.600	1.7532	1.7461	1.7403**				
5.800	1.7863	1.7813	1.7734**				
6.000	1.8184	1.8158	1.8055**				
6.200	1.8496	1.8496*	1.8366**				
* $\bar{U}(z)$		** $\bar{U}(z) - 0.0130$		* $\bar{U}(z)$		** $\bar{U}(z) - 0.0134$	

$K = 2.0$				$K = 5.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.	0.	0.	0.	0.
0.100	0.1987	0.1312	0.1312	0.100	0.4812	0.2375	0.2375
0.200	0.3900	0.2618	0.2617	0.200	0.8814	0.4724	0.4724
0.300	0.5688	0.3910	0.3909	0.300	1.1948	0.7022	0.7022
0.400	0.7327	0.5182	0.5181	0.400	1.4436	0.9246	0.9246
0.500	0.8814	0.6428	0.6427	0.500	1.6472	1.1377	1.1377
0.600	1.0160	0.7643	0.7642	0.600	1.8184	1.3399	1.3398
0.700	1.1380	0.8822	0.8822	0.700	1.9657	1.5298	1.5297
0.800	1.2490	0.9962	0.9961	0.800	2.0947	1.7068	1.7067
0.900	1.3504	1.1058	1.1057	0.900	2.2093	1.8705	1.8704
1.000	1.4436	1.2109	1.2108	1.000	2.3124	2.0208	2.0207
1.100	1.5297	1.3114	1.3112	1.100	2.4061	2.1583	2.1582
1.200	1.6094	1.4070	1.4068	1.200	2.4918	2.2835	2.2833
1.300	1.6837	1.4978	1.4976	1.300	2.5708	2.3972	2.3970
1.400	1.7532	1.5839	1.5836	1.400	2.6441	2.5005	2.5002
1.500	1.8184	1.6653	1.6650	1.500	2.7125	2.5943	2.5939
1.600	1.8799	1.7422	1.7419	1.600	2.7765	2.6796	2.6792
1.700	1.9379	1.8148	1.8144	1.700	2.8367	2.7576	2.7569
1.800	1.9928	1.8832	1.8828	1.800	2.8934	2.8289	2.8281
1.900	2.0450	1.9478	1.9473	1.900	2.9472	2.8947	2.8936
2.000	2.0947	2.0088	2.0082	2.000	2.9982	2.9555	2.9541
2.100	2.1421	2.0663	2.0656	2.100	3.0468	3.0121	3.0102
2.200	2.1874	2.1207	2.1198	2.200	3.0931	3.0651	3.0625
2.300	2.2308	2.1721	2.1711	2.300	3.1374	3.1151	3.1115

TABLE II — *Continued*

$K = 2.0$				$K = 5.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
2.400	2.2724	2.2209	2.2197	2.400	3.1798	3.1625	3.1576
2.500	2.3124	2.2672	2.2657	2.500	3.2205	3.2078	3.2010
2.600	2.3509	2.3113	2.3094	2.600	3.2596	3.2516	3.2420
2.700	2.3880	2.3533	2.3509	2.700	3.2972	3.2943	3.2807
2.800	2.4238	2.3934	2.3905	2.800	3.3335	3.3335*	3.3172**
2.900	2.4584	2.4318	2.4282				
3.000	2.4918	2.4687	2.4641				
3.100	2.5241	2.5042	2.4984				
3.200	2.5555	2.5384	2.5310				
3.300	2.5859	2.5716	2.5621				
3.400	2.6154	2.6038	2.5917**				
3.500	2.6441	2.6352	2.6204**				
3.600	2.6720	2.6660	2.6483**				
3.700	2.6992	2.6964	2.6755**				
3.800	2.7256	2.7256*	2.7019**				
* $\bar{U}(z)$		** $\bar{U}(z) - 0.0237$		* $\bar{U}(z)$		** $\bar{U}(z) - 0.0163$	

$K = 10.0$				$K = 20.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.	0.	0.	0.	0.
0.100	0.8814	0.3549	0.3549	0.100	1.4436	0.5146	0.5146
0.200	1.4436	0.7035	0.7035	0.200	2.0947	1.0147	1.0147
0.300	1.8184	1.0398	1.0397	0.300	2.4918	1.4870	1.4870
0.400	2.0947	1.3585	1.3585	0.400	2.7765	1.9206	1.9205
0.500	2.3124	1.6555	1.6554	0.500	2.9982	2.3079	2.3078
0.600	2.4918	1.9278	1.9278	0.600	3.1798	2.6453	2.6452
0.700	2.6441	2.1739	2.1739	0.700	3.3335	2.9330	2.9329
0.800	2.7765	2.3935	2.3934	0.800	3.4667	3.1746	3.1745
0.900	2.8934	2.5875	2.5874	0.900	3.5843	3.3756	3.3755
1.000	2.9982	2.7576	2.7575	1.000	3.6895	3.5427	3.5424
1.100	3.0931	2.9063	2.9060	1.100	3.7847	3.6823	3.6820
1.200	3.1798	3.0361	3.0358	1.200	3.8716	3.8005	3.8000
1.300	3.2596	3.1497	3.1494	1.300	3.9516	3.9020	3.9012
1.400	3.3335	3.2498	3.2493	1.400	4.0257	3.9910	3.9896
1.500	3.4023	3.3386	3.3379	1.500	4.0946	4.0702	4.0680
1.600	3.4667	3.4181	3.4171	1.600	4.1591	4.1423	4.1385
1.700	3.5272	3.4900	3.4885	1.700	4.2197	4.2089	4.2024
1.800	3.5843	3.5556	3.5534	1.800	4.2769	4.2719	4.2604**
1.900	3.6383	3.6161	3.6128	1.900	4.3309	4.3309*	4.3145**
2.000	3.6895	3.6723	3.6673				
2.100	3.7382	3.7252	3.7175				
2.200	3.7847	3.7754	3.7640**				
2.300	3.8291	3.8235	3.8084**				
2.400	3.8716	3.8704	3.8509**				
2.500	3.9124	3.9124*	3.8917**				
* $\bar{U}(z)$		** $\bar{U}(z) - 0.0207$		* $\bar{U}(z)$		** $\bar{U}(z) - 0.0164$	

TABLE II — *Continued*

$K = 50.0$				$K = 100.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.	0.	0.	0.	0.
0.050	1.6472	0.4086	0.4086	0.050	2.3124	0.5671	0.5671
0.100	2.3124	0.8120	0.8120	0.100	2.9982	1.1232	1.1231
0.150	2.7125	1.2052	1.2052	0.150	3.4023	1.6579	1.6578
0.200	2.9982	1.5835	1.5835	0.200	3.6895	2.1615	2.1613
0.250	3.2205	1.9427	1.9427	0.250	3.9124	2.6260	2.6257
0.300	3.4023	2.2793	2.2792	0.300	4.0946	3.0454	3.0451
0.350	3.5562	2.5905	2.5904	0.350	4.2487	3.4162	3.4158
0.400	3.6895	2.8746	2.8745	0.400	4.3822	3.7378	3.7373
0.450	3.8072	3.1309	3.1308	0.450	4.4999	4.0120	4.0114
0.500	3.9124	3.3595	3.3595	0.500	4.6053	4.2429	4.2422
0.550	4.0077	3.5617	3.5616	0.550	4.7006	4.4358	4.4349
0.600	4.0946	3.7391	3.7390	0.600	4.7876	4.5968	4.5956
0.650	4.1746	3.8941	3.8940	0.650	4.8676	4.7316	4.7299
0.700	4.2487	4.0293	4.0291	0.700	4.9417	4.8458	4.8433
0.750	4.3177	4.1471	4.1470	0.750	5.0107	4.9438	4.9403
0.800	4.3822	4.2503	4.2501	0.800	5.0752	5.0297	5.0244
0.850	4.4428	4.3411	4.3408	0.850	5.1358	5.1066	5.0986
0.900	4.4999	4.4215	4.4211	0.900	5.1930	5.1774	5.1649
0.950	4.5540	4.4935	4.4930	0.950	5.2471	5.2471*	5.2250
1.000	4.6053	4.5584	4.5577	1.000	5.2983	5.2983*	5.2799
1.050	4.6541	4.6176	4.6167	1.050	5.3471	5.3471*	5.3301
1.100	4.7006	4.6721	4.6708	1.100	5.3936	5.3936*	5.3767**
1.150	4.7450	4.7228	4.7209				
1.200	4.7876	4.7703	4.7675				
1.250	4.8284	4.8152	4.8111				
1.300	4.8676	4.8580	4.8520				
1.350	4.9053	4.8993	4.8904				
1.400	4.9417	4.9396	4.9268**				
1.450	4.9768	4.9768*	4.9619**				
* $\bar{U}(z)$		** $\bar{U}(z) - 0.0149$		* $\bar{U}(z)$		** $\bar{U}(z) - 0.0170$	

$K = 200.0$				$K = 500.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.	0.	0.	0.	0.
0.020	2.0947	0.3120	0.3120	0.020	2.9982	0.4669	0.4669
0.040	2.7765	0.6226	0.6226	0.040	3.6895	0.9300	0.9300
0.060	3.1798	0.9302	0.9302	0.060	4.0946	1.3856	1.3855
0.080	3.4667	1.2335	1.2334	0.080	4.3822	1.8299	1.8298
0.100	3.6895	1.5310	1.5309	0.100	4.6053	2.2595	2.2594
0.120	3.8716	1.8214	1.8212	0.120	4.7876	2.6709	2.6708
0.140	4.0257	2.1034	2.1032	0.140	4.9417	3.0613	3.0612
0.160	4.1591	2.3758	2.3756	0.160	5.0752	3.4280	3.4278
0.180	4.2769	2.6375	2.6373	0.180	5.1930	3.7689	3.7687
0.200	4.3822	2.8877	2.8875	0.200	5.2983	4.0825	4.0823
0.220	4.4775	3.1254	3.1252	0.220	5.3936	4.3680	4.3678
0.240	4.5645	3.3501	3.3499	0.240	5.4807	4.6254	4.6251
0.260	4.6445	3.5613	3.5611	0.260	5.5607	4.8552	4.8549
0.280	4.7186	3.7588	3.7585	0.280	5.6348	5.0588	5.0584

TABLE II — *Continued*

$K = 200.0$				$K = 500.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.300	4.7876	3.9425	3.9421	0.300	5.7038	5.2378	5.2374
0.320	4.8521	4.1125	4.1121	0.320	5.7683	5.3946	5.3941
0.340	4.9127	4.2691	4.2687	0.340	5.8290	5.5313	5.5308
0.360	4.9699	4.4128	4.4123	0.360	5.8861	5.6506	5.6499
0.380	5.0239	4.5442	4.5437	0.380	5.9402	5.7547	5.7539
0.400	5.0752	4.6640	4.6635	0.400	5.9915	5.8460	5.8450
0.420	5.1240	4.7731	4.7725	0.420	6.0403	5.9264	5.9251
0.440	5.1705	4.8722	4.8715	0.440	6.0868	5.9978	5.9961
0.460	5.2150	4.9622	4.9614	0.460	6.1312	6.0616	6.0594
0.480	5.2575	5.0440	5.0431	0.480	6.1738	6.1192	6.1163
0.500	5.2983	5.1185	5.1174	0.500	6.2146	6.1718	6.1678
0.520	5.3376	5.1863	5.1850	0.520	6.2538	6.2201	6.2147
0.540	5.3753	5.2483	5.2467	0.540	6.2916	6.2649	6.2577
0.560	5.4117	5.3051	5.3033	0.560	6.3279	6.3069	6.2970
0.580	5.4468	5.3574	5.3552	0.580	6.3630	6.3466	6.3329**
0.600	5.4807	5.4057	5.4030	0.600	6.3969	6.3843	6.3668**
0.620	5.5134	5.4506	5.4473	0.620	6.4297	6.4206	6.3996**
0.640	5.5452	5.4924	5.4884	0.640	6.4615	6.4557	6.4313**
0.660	5.5760	5.5316	5.5267	0.660	6.4922	6.4902	6.4621**
0.680	5.6058	5.5685	5.5624	0.680	6.5221	6.5221*	6.4920**
0.700	5.6348	5.6034	5.5958				
0.720	5.6630	5.6365	5.6270				
0.740	5.6904	5.6682	5.6562				
0.760	5.7170	5.6986	5.6834**				
0.780	5.7430	5.7278	5.7094**				
0.800	5.7683	5.7561	5.7347**				
0.820	5.7930	5.7836	5.7594**				
0.840	5.8171	5.8105	5.7835**				
0.860	5.8406	5.8370	5.8070**				
0.880	5.8636	5.8636*	5.8300**				
* $\bar{U}(z)$		** $\bar{U}(z) - 0.0336$		* $\bar{U}(z)$		** $\bar{U}(z) - 0.0301$	

$K = 1000.0$				$K = 2000.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.	0.	0.	0.	0.
0.020	3.6895	0.6277	0.6277	0.010	3.6895	0.4202	0.4202
0.040	4.3822	1.2477	1.2476	0.020	4.3822	0.8384	0.8384
0.060	4.7876	1.8523	1.8522	0.030	4.7876	1.2527	1.2526
0.080	5.0752	2.4342	2.4341	0.040	5.0752	1.6611	1.6611
0.100	5.2983	2.9864	2.9863	0.050	5.2983	2.0619	2.0618
0.120	5.4807	3.5027	3.5025	0.060	5.4807	2.4530	2.4529
0.140	5.6348	3.9777	3.9775	0.070	5.6348	2.8327	2.8326
0.160	5.7683	4.4075	4.4073	0.080	5.7683	3.1992	3.1991
0.180	5.8861	4.7899	4.7897	0.090	5.8861	3.5510	3.5509
0.200	5.9915	5.1245	5.1243	0.100	5.9915	3.8866	3.8865
0.220	6.0868	5.4128	5.4126	0.110	6.0868	4.2046	4.2044
0.240	6.1738	5.6582	5.6579	0.120	6.1738	4.5039	4.5038
0.260	6.2538	5.8649	5.8645	0.130	6.2538	4.7838	4.7837
0.280	6.3279	6.0381	6.0376	0.140	6.3279	5.0438	5.0436
0.300	6.3969	6.1831	6.1825	0.150	6.3969	5.2834	5.2833
0.320	6.4615	6.3049	6.3041	0.160	6.4615	5.5030	5.5028

TABLE II — *Continued*

$K = 1000.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.340	6.5221	6.4082	6.4071
0.360	6.5793	6.4967	6.4952
0.380	6.6333	6.5738	6.5716
0.400	6.6846	6.6420	6.6390
0.420	6.7334	6.7036	6.6991
0.440	6.7799	6.7601	6.7535
0.460	6.8244	6.8133	6.8033
0.480	6.8669	6.8644	6.8493
0.500	6.9078	6.9078*	6.8919
0.520	6.9470	6.9470*	6.9315**

* $\bar{U}(z)$ ** $\bar{U}(z) - 0.0155$

$K = 2000.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.170	6.5221	5.7029	5.7027
0.180	6.5793	5.8838	5.8836
0.190	6.6333	6.0466	6.0464
0.200	6.6846	6.1926	6.1924
0.210	6.7334	6.3231	6.3228
0.220	6.7799	6.4394	6.4391
0.230	6.8244	6.5430	6.5426
0.240	6.8669	6.6354	6.6349
0.250	6.9078	6.7178	6.7172
0.260	6.9470	6.7916	6.7909
0.270	6.9847	6.8578	6.8570
0.280	7.0211	6.9176	6.9166
0.290	7.0562	6.9719	6.9707
0.300	7.0901	7.0215	7.0200
0.310	7.1229	7.0672	7.0653
0.320	7.1546	7.1094	7.1071
0.330	7.1854	7.1489	7.1458
0.340	7.2152	7.1859	7.1821
0.350	7.2442	7.2210	7.2161
0.360	7.2724	7.2545	7.2481
0.370	7.2998	7.2867	7.2785
0.380	7.3265	7.3179	7.3072
0.390	7.3524	7.3485	7.3345
0.400	7.3778	7.3778*	7.3604
0.410	7.4025	7.4025*	7.3851**

* $\bar{U}(z)$ ** $\bar{U}(z) - 0.0173$

$K = 5000.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.
0.010	4.6053	0.6118	0.6118
0.020	5.2983	1.2187	1.2187
0.030	5.7038	1.8157	1.8157
0.040	5.9915	2.3981	2.3981
0.050	6.2146	2.9610	2.9610
0.060	6.3969	3.4999	3.4998
0.070	6.5511	4.0105	4.0104
0.080	6.6846	4.4889	4.4888
0.090	6.8024	4.9318	4.9317
0.100	6.9078	5.3367	5.3366
0.110	7.0031	5.7021	5.7020
0.120	7.0901	6.0276	6.0275
0.130	7.1701	6.3139	6.3137
0.140	7.2442	6.5629	6.5627
0.150	7.3132	6.7774	6.7772
0.160	7.3778	6.9607	6.9605
0.170	7.4384	7.1169	7.1166
0.180	7.4955	7.2496	7.2492
0.190	7.5496	7.3626	7.3622

$K = 10000.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.	0.	0.	0.
0.005	4.6053	0.4051	0.4050
0.010	5.2983	0.8089	0.8089
0.015	5.7038	1.2102	1.2102
0.020	5.9915	1.6078	1.6078
0.025	6.2146	2.0005	2.0005
0.030	6.3969	2.3870	2.3870
0.035	6.5511	2.7662	2.7662
0.040	6.6846	3.1368	3.1368
0.045	6.8024	3.4977	3.4977
0.050	6.9078	3.8478	3.8477
0.055	7.0031	4.1860	4.1859
0.060	7.0901	4.5112	4.5111
0.065	7.1701	4.8226	4.8225
0.070	7.2442	5.1193	5.1192
0.075	7.3132	5.4006	5.4005
0.080	7.3778	5.6659	5.6658
0.085	7.4384	5.9149	5.9148
0.090	7.4955	6.1472	6.1472
0.095	7.5496	6.3630	6.3629

TABLE II — *Concluded*

$K = 5000.0$				$K = 10000.0$			
z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$	z	$\bar{U}(z)$	$U_+(z)$	$U_-(z)$
0.200	7.6009	7.4595	7.4589	0.100	7.6009	6.5622	6.5621
0.210	7.6497	7.5431	7.5424	0.105	7.6497	6.7453	6.7452
0.220	7.6962	7.6161	7.6151	0.110	7.6962	6.9128	6.9126
0.230	7.7407	7.6805	7.6792	0.115	7.7407	7.0653	7.0652
0.240	7.7832	7.7382	7.7364	0.120	7.7832	7.2038	7.2036
0.250	7.8240	7.7907	7.7882	0.125	7.8240	7.3290	7.3289
0.260	7.8633	7.8390	7.8355	0.130	7.8633	7.4421	7.4419
0.270	7.9010	7.8841	7.8792	0.135	7.9010	7.5440	7.5438
0.280	7.9374	7.9270	7.9200	0.140	7.9374	7.6358	7.6355
0.290	7.9725	7.9685	7.9583	0.145	7.9725	7.7184	7.7181
0.300	8.0064	8.0064*	7.9947	0.150	8.0064	7.7929	7.7926
0.310	8.0392	8.0392*	8.0292	0.155	8.0392	7.8602	7.8598
0.320	8.0709	8.0709*	8.0623	1.160	8.0709	7.9212	7.9207
0.330	8.1017	8.1017*	8.0939	0.165	8.1017	7.9765	7.9760
0.340	8.1315	8.1315*	8.1243	0.170	8.1315	8.0270	8.0264
0.350	8.1605	8.1605*	8.1535**	0.175	8.1605	8.0733	8.0726
				0.180	8.1887	8.1159	8.1151
				0.185	8.2161	8.1554	8.1543
				0.190	8.2428	8.1921	8.1908
				0.195	8.2687	8.2265	8.2249
				0.200	8.2940	8.2589	8.2569
				0.205	8.3187	8.2896	8.2871
				0.210	8.3428	8.3187	8.3157
				0.215	8.3664	8.3466	8.3429
				0.220	8.3894	8.3735	8.3688
				0.225	8.4118	8.3995	8.3936
				0.230	8.4338	8.4248	8.4174
				0.235	8.4553	8.4496	8.4402
				0.240	8.4764	8.4740	8.4622
				0.245	8.4970	8.4970*	8.4833
				0.250	8.5172	8.5172*	8.5036**
* $\bar{U}(z)$	** $\bar{U}(z) - 0.0070$			* $\bar{U}(z)$	** $\bar{U}(z) - 0.0135$		

They have been rounded off and retabulated from the original IBM results, which were at shorter intervals and nominally to eight significant figures.

It goes without saying that one must be wary of roundoff and truncation errors when trying to integrate a stiff differential equation numerically. Even if one could start with exactly the right initial conditions, accumulated errors would shift the computation to a neighboring integral curve which would ultimately diverge rapidly from the desired solution. Strictly speaking, we cannot even be sure of the upper and lower bounds computed by stepwise integration of the differential equation, since roundoff errors could make two numerical solutions with nearly equal initial slopes cross each other and tend to infinity in the wrong directions. (No such crossing was observed in the present study.)

Unfortunately it is not easy, especially for the Runge-Kutta method, to obtain error bounds which are both rigorous and realistic. We therefore fell back on the usual test which is made to judge whether the error in a numerical integration is tolerable; namely, we ran the integration again with a double interval and compared results. This test was applied to most of the upper and lower bounds included in the tables. The difference between the two integrations was found to increase rapidly toward the end of the range, but always to be small compared to the difference between upper and lower bounds. In the worst case ($K = 0.1$) the difference between the two integrations for a given bound was less than one-tenth of the difference between upper and lower bounds; and for larger K the discrepancies were considerably smaller. We therefore feel quite confident that the tabulated results really are upper and lower bounds for the exact solution.

The range of the calculations shown here was chosen to bridge the gap between available approximations for small K (the "neutral" case) and large K (the "space-charge" case). For $K < 0.1$ the present method begins to run into instability trouble, though it could probably be modified to work for smaller values of K in the unlikely event that such calculations are wanted. For large K, on the other hand, there is no indication that $K = 10{,}000$ marks the upper limit of what can be handled with this technique.

In summary, then, a program has been written for the IBM 704 which will produce in a few minutes, for any value of K between 0.1 and somewhere above 10,000, very close upper and lower bounds for the potential distribution in a linearly graded p-n junction. A minor modification of the program would permit similar calculations for a junction with any reasonable distribution of fixed charge. So far as numerical techniques are concerned, therefore, the problem may be regarded as completely solved.

APPENDIX B

Numerical Capacitance Calculations

The neutral capacitance of a junction of normalized width z_l is given by (23) of the text as

$$C_n = \frac{C_0}{K^{\frac{1}{3}}}\left[I(z_l, K) - 3\frac{\partial I(z_l, K)}{\partial(\log K)} + z_l e^{-U_l}\right], \tag{B1}$$

where

$$I(z_l, K) = \int_0^{z_l} e^{-U(z,K)}\, dz. \tag{B2}$$

The value of I may be calculated numerically from Table II, which gives $U(z, K)$ for 16 values of K ranging from 0.1 to 10,000. Extension to arbitrarily large values of z_l is straightforward, since for sufficiently large z the tabulated functions differ by at most a constant from the asymptote $\bar{U}(z, K) = \sinh^{-1}Kz$. If U is replaced by $\bar{U} - \epsilon$ for $z \geqq z_0 \geqq 0$, where ϵ is constant, then provided that $z_l \geqq z_0$ and U is continuous at z_0, the expression for $I(z_l, K)$ becomes

$$I(z_l, K) = \int_0^{z_0} e^{-U} dz + \frac{e^{\epsilon}(U_l - U_0)}{2K} + \frac{z_l e^{-U_l} - z_0 e^{-U_0}}{2}. \quad \text{(B3)}$$

When K is small (the neutral case), then $U \approx \bar{U}$ for all z, and (B1) and (B3) yield the approximation (26).

The numerical evaluation of C_n/C_0 was carried out by evaluating $I(z_l, K)$ for each value of K using Simpson's rule. The function $\log I$ was then fitted at three adjacent values of K by a quadratic polynomial in $\log K$, and the derivative of I was found approximately by differentiating the polynomial at the middle value of K (except, of course, at $K = 0.1$ and $K = 10{,}000$, where the end value had to be used). The value of z_l was chosen to make $\bar{U}(z_l, K_1) = 10$, where K_1 is the point at which the derivative is evaluated. The value of z_l is supposed to be held constant during the differentiation.

Table III shows the values found for C_n/C_0 as a function of K. These values are actually averages of the results found by taking for $U(z, K)$ the upper and lower bounds given in Table II. The two numbers averaged differed from each other by less than 2 per cent in all cases, and usually by less than 1 per cent. As is well known, numerical differentia-

TABLE III — NORMALIZED CAPACITANCES

K	C_n/C_0	C_p/C_0
0.1	432	0.423
0.2	171	0.608
0.5	50.5	0.889
1	20.1	1.052
2	8.05	1.139
5	2.44	1.156
10	1.02	1.123
20	0.436	1.082
50	0.152	1.020
100	0.0721	0.972
200	0.0360	0.931
500	0.0153	0.883
1000	0.00834	0.850
2000	0.00468	0.822
5000	0.00225	0.790
10000	0.00132	0.762

tion magnifies the errors in inaccurate data. One may hope, however, that the values obtained for C_n are accurate to about 1 per cent. Comparison with the approximate formula (26) is shown in Fig. 6. Agreement is good for $K \lesssim 5$, but the approximation gives too low a value for C_n when K is large.

The space-charge capacitance is given by (24) as

$$C_\rho = \frac{C_0}{K^{\frac{1}{2}}}\left[3\,\frac{\partial S}{\partial(\log K)} - S\right], \tag{B4}$$

where

$$S = U'(0, K) \tag{B5}$$

is the initial slope of the potential function. The initial slope is given, as a function of K, to five significant figures in Table I. Logarithmic differentiation, using a three-point interpolating function, leads to the values of C_ρ/C_0 shown in Table III (the difference between the upper and lower bounds of the slope rarely affects the third decimal place). Again the accuracy is not precisely known, but is presumed to be 1 per cent or better.

For the neutral case, say $K \lesssim 0.2$, one has $S \approx K$ and

$$C_\rho \approx 2C_0K^{\frac{1}{2}}. \tag{B6}$$

When $K \gg 1$, the initial slope according to the space-charge approximation is

$$s = \left(\frac{3U_m \sinh U_m}{4}\right)^{\frac{1}{2}}, \tag{B7}$$

where U_m is defined in terms of K by (28). A straightforward if somewhat laborious substitution of (B7) and (28) into (B4) leads to

$$C_\rho = \frac{2(3U_m)^{\frac{1}{2}}C_0}{3U_m \coth U_m - 1}, \tag{B8}$$

as in (27). The approximation to C_ρ/C_0 given by (B8) is plotted against K in Fig. 6. It is seen that the approximate value of C_ρ lies above the correct value for K less than about 6, and below it, by amounts ranging up to nearly 10 per cent, for larger K.

REFERENCES

1. Shockley, W., Theory of p-n Junctions in Semiconductors and p-n Junction Transistors, B.S.T.J., **28**, 1949, p. 435; especially Section 2.
2. Moll, J. L., unpublished work.
3. Prim, R. C., unpublished work.
4. Morrison, J. A., unpublished work.
5. Kamke, E., *Differentialgleichungen*, 3rd ed., Chelsea, New York, 1948, p. 150.

Transverse Electron Beam Waves in Varying Magnetic Fields

By E. I. GORDON

(Manuscript received June 23, 1960)

The properties of electron cyclotron and synchronous waves in varying magnetic fields are discussed. Magnetic field variations in space and time are considered. The problem is treated by establishing the wave excitation from knowledge of the macroscopic beam motion. It is shown that the cyclotron wave is coupled to the synchronous wave and that both waves are always amplified in a changing field. Unless the charge density is an appreciable fraction of the full Brillouin value, however, the individual electron orbits will be amplified along with the waves causing beam expansion.

The phase velocity of the waves is shown to be approximately independent of space charge. In the case of a spatially varying field, one of the waves must be fast, carrying positive kinetic power, and the other slow, carrying negative kinetic power. The total kinetic power carried by the two waves is conserved. When the magnetic field varies in time, the kinetic power of the two waves is not conserved but the Manley-Rowe relation is satisfied. When the field varies at a rate greater than the signal frequency, both modes may carry positive kinetic power.

I. INTRODUCTION

The electron cyclotron wave has received a great deal of attention in the past year, since it forms the basis of a successful beam-type traveling-wave parametric amplifier. As was first demonstrated by Adler et al.,[1,2,3] the fast cyclotron wave is actively coupled to a fast cyclotron wave idler by the action of a high-frequency transverse electric quadrupole field. It was shown by Gordon[4,5] that space periodic quadrupole fields actively couple a fast and slow cyclotron wave.

The purpose of this paper is to demonstrate that cyclotron waves will also be amplified in varying magnetic fields. It will be shown that the idler mode associated with the cyclotron wave is a synchronous wave associated with the spatial configuration of the beam. Analogous to the quadrupole case, pumping in a time-varying magnetic field with the

pump frequency greater than the signal frequency leads to active coupling of two fast modes. When the pump frequency is lower than the signal frequency one mode must be fast and the other slow. Pumping in a spatially varying magnetic field will be seen to correspond to the limiting case of zero pump frequency.

In Section II the cyclotron and synchronous waves will be described and some simple points of view established. Then the results of an earlier analysis,[6] in which a general solution for charged particle orbits in arbitrarily varying magnetic fields is given, will be utilized to determine the properties of coupled cyclotron and synchronous waves. The analysis is done in such a way as to include the possibility of pumping with time- and space-varying fields and succeeding sections will be devoted to each. Space-charge effects will be shown to be significant in preventing beam expansion.

II. CYCLOTRON AND SYNCHRONOUS WAVES

A wave description of the motion of an electron beam is appropriate only to the extent that it can describe the energy exchange between the beam and the electromagnetic field. In the case of transverse beam waves, the beam may be thought of as made up of thin discs with no longitudinal coupling. The appropriate wave variable is the center-of-mass of the disc. The validity of this model is justified by considering the significant coupling term between the electromagnetic field and the beam, namely the time average value of $\int \mathbf{J}\cdot\mathbf{E}d^3r$, in which $\mathbf{J}$ is the electron conduction current and $\mathbf{E}$ is the RF electric field. In a given unit-volume cross section of the beam, the significant quantities which contribute to the time average value are

$$\mathbf{J}\cdot\mathbf{E} = -e\sum_{j=1}^{N}(\dot{\mathbf{r}}_j\cdot\mathbf{E}_r + v_j\mathbf{r}_j\cdot\text{grad}_r E_z),$$

in which N is the electron density, v_j is the z-directed drift velocity and $\mathbf{r}_j$ is the transverse coordinate of the jth electron. It is assumed that the z-axis of the coordinate system is at a maximum of the transverse field and a zero of the longitudinal field. For a thin beam placed along the z-axis, $\mathbf{E}_r$ and $\text{grad}_r E_z$ are constant over the cross section of the beam to within terms that are of order $(\beta r_j)^2 \ll 1$, in which β is the wave propagation constant. To the same degree of approximation, variations in the drift velocity resulting from longitudinal forces also may be neglected and v_j may be assumed to be equal to the unperturbed drift velocity.

Hence, the quantities

$$\mathbf{R} = N^{-1} \sum_{j=1}^{N} \mathbf{r}_j \qquad \text{and} \qquad \dot{\mathbf{R}} = N^{-1} \sum_{j=1}^{N} \dot{\mathbf{r}}_j ,$$

i.e., the position and velocity of the center-of-mass of the disc, are sufficient to describe the interaction. Since the external transverse forces on each electron in the cross section are identical and the internal forces between electrons cancel exactly, the motion of the disc is exactly the same as that of a single electron and is independent of the internal motions of the electrons within the disc or the beam density. Hence, we may conclude that the transverse waves are independent of the beam space charge.

In the uniform coupler fields the internal motion of the electrons with respect to the center-of-mass proceeds as if no external forces were applied. In pump fields the significant forces vary over the cross section of the beam. As a result the individual electron trajectories measured in the center-of-mass coordinate system may also be pumped. This can be avoided by making the beam sufficiently dense that the internal motions of the electrons within the beam are sufficiently different from the motion of the center-of-mass itself. This point will be discussed in detail later.

In the earlier paper,[6] it is shown that the position of a charged particle in a plane transverse to the applied magnetic field is conveniently described in terms of a complex vector quantity $\mathbf{r} = x + iy = \mathbf{r}_g + \mathbf{r}_b$, in which x and y are the rectangular coordinates of the charged particle in the transverse plane. The quantity $\mathbf{r}_g$ represents the instantaneous center of curvature of the charged particle orbit, called the guiding center, and $\mathbf{r}_b$, called the radius vector, denotes the instantaneous radius of curvature of the trajectory and the position of the particle with respect to the guiding center (see Fig. 1). In a uniform magnetic field,

$$\mathbf{r} = \mathbf{r}_{g0} + \mathbf{r}_{b0} e^{i\omega_c (t - t_0)}, \tag{1}$$

in which $\mathbf{r}_{g0}$ and $\mathbf{r}_{b0}$ are the appropriate values at $t = t_0$ and ω_c is the cyclotron frequency. The motion of the beam disc is precisely the same. The four independent quantities, x, $\dot{x}$, y and $\dot{y}$, of the beam disc center-of-mass make possible four transverse waves on the electron beam. These are the fast and slow cyclotron and synchronous waves. It will be seen that the radius vector and guiding center of the beam disc are the appropriate quantities to express the cyclotron and synchronous wave excitation of the beam.

The cyclotron wave is excited by a circularly polarized transverse

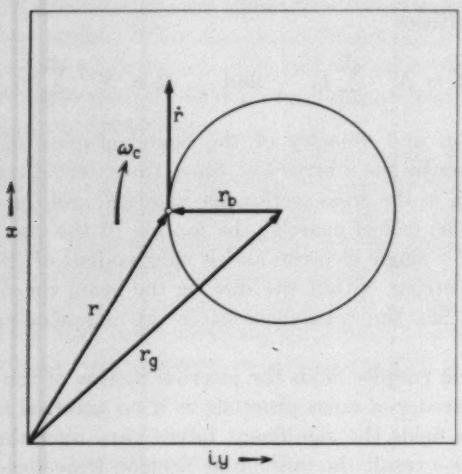


Fig. 1 — Particle coordinates for motion in a magnetic field directed along the z-axis.

electric field. The electric vector may rotate in the same or the opposite direction to the natural cyclotron motion of the disc. If the exciting electric field has a signal component at a frequency ω_s then we may say that ω_s is positive if the electric field rotates with the disc and is negative if the electric field rotates in the opposite direction (see Fig. 2). The initial phase of the field-excited rotation of the disc depends on the phase of the field at the time of entry. As a result, a cyclotron wave excitation at a signal frequency ω_s can be described in the following way: Each disc rotates about the axis with an angular velocity ω_c. An observer, standing at a fixed plane, will note that the *phase angle of rotation* of successive discs as they pass through the plane *will change at a rate* ω_s because of the spatial twist of the beam. If ω_s is positive, the excitation is a fast cyclotron wave. The slow wave is associated with a negative value of ω_s. If a cyclotron wave excitation exists, the phase angle of rotation of any disc which passes some plane, say $z = 0$, at a time $t = t_0$ can always be written as $\omega_s t_0$. The radius vector of the rotational motion can then be written

$$\mathbf{R}_{b0} = R_{b0} e^{i\omega_s t_0}, \tag{2}$$

in which R_{b0} is a real number representing the radius of the disc orbit. Thus the radius vector of any disc along the beam, with arbitrary t_0, can be written for a uniform static field as

$$\mathbf{R}_b = \mathbf{R}_{b0} e^{i\omega_c(t-t_0)} = R_{b0} e^{i\omega_s(t-z/v_p)}, \tag{3}$$

in which we have used the relation, $t - t_0 = z/v$, v being the electron drift velocity. Clearly this is a wave motion for which $v_p = \omega_s v/(\omega_s - \omega_c)$ is the phase velocity. Note that when ω_s is positive, v_p is either greater than v or negative, and the wave is said to be "fast." When ω_s is negative, v_p is always less than v and the wave is said to be "slow."

In Fig. 2 the cyclotron wave is assumed to be in synchronism with a circularly polarized electromagnetic field of phase velocity u. The Doppler-shifted frequency in the frame of reference moving with the beam is $\omega_s(1 - v/u)$. Equating $v_p = u$ yields $\omega_c = \omega_s(1 - v/u)$; hence, synchronism implies that cyclotron resonance occurs at the Doppler-shifted frequency and the rotating disc remains in phase with the rotating uniform electric field, $E_\perp$.

The longitudinal forces on the rotating disc arise from the uniform transverse magnetic field, $H_\perp$, and longitudinal electric field, $E_\parallel$, which increases with distance from the axis. The forces excited by these fields are proportional to the rotational velocity and radius respectively. Since the electromagnetic fields in free space are related by Maxwell's equations, and the transverse velocity and radius are related by the cyclotron frequency, the changes in rotational and drift energy of each disc are uniquely related independent of the structure propagating the electromagnetic wave.

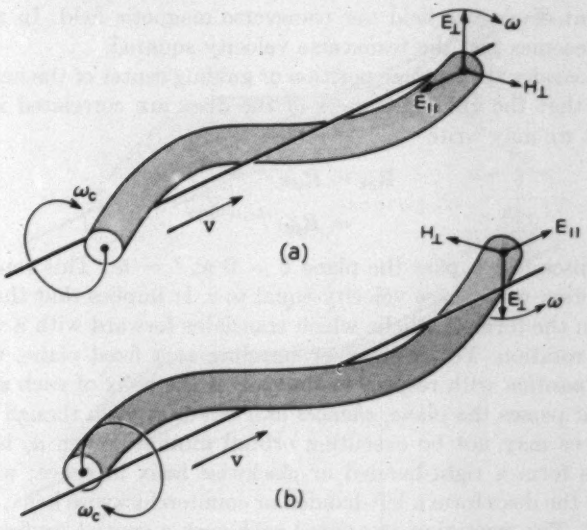


Fig. 2 — Beam motion for (a) fast and (b) slow cyclotron waves excited at a frequency ω.

Each disc experiences precisely the same forces and hence undergoes the same changes in rotational and drift energy. As a result, the total power carried by the beam changes. The power change is called the kinetic power of the beam wave and for the cyclotron wave it has two components:

$$\begin{aligned} P_{k\perp} &= \tfrac{1}{2}m(I_0/e)\omega_c^2 R_b^2, \\ P_{k\parallel} &= \tfrac{1}{2}m(I_0/e)\omega_c(\omega_s - \omega_c)R_b^2, \end{aligned} \tag{4}$$

in which I_0 is the beam current. The total kinetic power $P_k = P_{k\perp} + P_{k\parallel}$ has the value[7,8,9]

$$P_k = \tfrac{1}{2}m(I_0/e)\omega_s\omega_c R_b^2. \tag{5}$$

The longitudinal kinetic power of the slow cyclotron wave is always negative ($\omega_s < 0$) and larger in magnitude than the transverse kinetic power. The longitudinal kinetic power of the fast cyclotron wave may be positive or negative depending upon whether the fast wave is a forward wave ($\omega_s > \omega_c$) or a backward wave ($\omega_s < \omega_c$). When it is negative its magnitude is always less than that of the transverse energy. When $\omega_s = \omega_c$, the longitudinal part of the kinetic power is zero since the phase velocity becomes infinite. Such a wave is excited by electromagnetic waves of infinite phase velocity which do not have a longitudinal component of electric field nor transverse magnetic field. In this limit $\omega_s\omega_c R_b^2$ becomes just the transverse velocity squared.

Now consider the average position or guiding center of the beam discs. Suppose that the guiding centers of the discs are correlated in such a way that we may write

$$\begin{aligned} \mathbf{R}_{g0} &= R_{g0}e^{-i\omega_s t_0} \\ &= R_{g0}e^{-i\omega_s(t-z/v)} \end{aligned} \tag{6}$$

for the discs which pass the plane $z = 0$ at $t = t_0$. This represents a wave motion with phase velocity equal to v. It implies that the beam is shaped in the form of a helix which translates forward with a velocity v without rotation. To an observer standing at a fixed plane, the *phase angle of position* with respect to the axis of the helix of each successive disc, as it passes the plane, *changes at a rate* $-\omega_s$ even though the discs themselves may not be executing orbital motion. When ω_s is positive the discs form a right-handed or clockwise helix in space; when ω_s is negative the discs form a left-handed or counterclockwise helix, as shown in Fig. 3. The excitation associated with such a spatial configuration is known as a synchronous wave.

The synchronous wave is also excited by circularly polarized electromagnetic fields. Synchronism requires $v = u$, so that the Doppler-shifted frequency as observed by the drifting discs is zero and they remain in phase with the exciting field always. Hence the discs will drift radially outward with a velocity $E_\perp/B$, in which B is the DC magnetic flux density, in a direction depending upon the phase of the field $E_\perp$. When the field is removed, the discs stop drifting outward and there is no net change in the transverse energy. The change in drift energy comes from the longitudinal electric field. The kinetic power is given by[7,8,9]

$$P_k = \tfrac{1}{2}m(I_0/e)\omega_s\omega_c R_g^2. \tag{7}$$

The synchronous wave which forms a left-handed helix carries negative kinetic power, while the wave which forms a right-handed helix carries positive kinetic power, as can be seen from Fig. 3. Consequently, a convention of positive frequencies referring to positive kinetic power waves and negative frequencies to negative kinetic power waves is pertinent to both cyclotron and synchronous waves. Henceforth, the terms "fast" and "slow" synchronous waves will be used in reference to their kinetic power, although the phase velocity of both waves is always synchronous with the drift velocity of the beam.

The kinetic power carried by the synchronous wave is proportional to

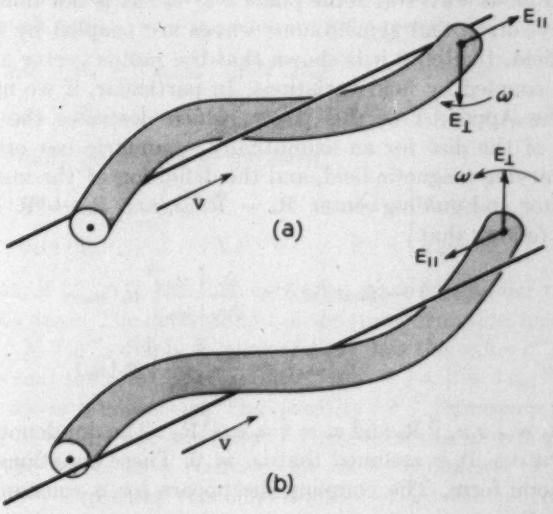


Fig. 3 — Beam motion for (b) fast and (a) slow synchronous waves excited at a frequency ω.

R_g^2. In this case the guiding center plays the role that the radius vector plays in the cyclotron wave. For either wave the important quantity is the transverse displacement associated with the disc as has been shown by Bobroff[7] [see Eq. (50)], and more recently by Siegman[8] and Klüver.[9]

III. AMPLIFICATION IN SPATIALLY VARYING MAGNETIC FIELDS

Suppose that the initial excitation of the beam at $z = z_0$ consists of a cyclotron wave of frequency ω_s with normalized amplitude $a_{c0}(\omega_s)$,

$$\begin{aligned} a_{c0}(\omega_s) &= |\, \omega_s\omega_{c0}\,|^{\frac{1}{2}} R_{b0} e^{i\omega_s t_0} \\ &= |\, \omega_s\omega_{c0}\,|^{\frac{1}{2}} \mathbf{R}_{b0} \end{aligned} \tag{8}$$

and a synchronous wave of frequency ω_m with amplitude $a_{s0}(\omega_m)$,

$$\begin{aligned} a_{s0}(\omega_m) &= |\, \omega_m\omega_{c0}\,|^{\frac{1}{2}} R_{g0} e^{-i\omega_s t_0} \\ &= |\, \omega_m\omega_{c0}\,|^{\frac{1}{2}} \mathbf{R}_{g0}\,. \end{aligned} \tag{9}$$

The phase of a_{c0} and a_{s0} at $z = z_0$ is included. The longitudinal magnetic field is assumed to be azimuthally symmetric but varying with both z and t. The cylotron frequency at $z = z_0$ has the value ω_{c0}. The purpose of the following analysis will be to find the amplitude of the cyclotron and synchronous waves at some plane $z > z_0$. It is not difficult to see that the cyclotron and synchronous waves are coupled by a changing magnetic field. In Ref. 6 it is shown that the radius vector and guiding center are coupled by field variations. In particular, if we make use of (29) of the Appendix of this paper, which describes the transverse trajectory of the disc for an azimuthally symmetric but otherwise arbitrarily varying magnetic field, and the definition of the instantaneous radius vector and guiding center, $\mathbf{R}_b = \dot{\mathbf{R}}/i\omega_c$ and $\mathbf{R}_g = \mathbf{R} - \mathbf{R}_b$. (see Fig. 1), it follows that

$$\begin{aligned} \dot{a}_c(\omega_s) - i\omega_c a_c(\omega_s) &= \frac{\dot{\omega}_c}{2\omega_c}\left|\frac{\omega_s}{\omega_m}\right|^{\frac{1}{2}} a_s(\omega_m), \\ \dot{a}_s(\omega_m) &= \frac{\dot{\omega}_c}{2\omega_c}\left|\frac{\omega_m}{\omega_s}\right|^{\frac{1}{2}} a_c(\omega_s), \end{aligned} \tag{10}$$

in which $a_c = |\, \omega_s\omega_c\,|^{\frac{1}{2}} \mathbf{R}_b$ and $a_s = |\, \omega_m\omega_c\,|^{\frac{1}{2}} \mathbf{R}_g$. The dot denotes the total time derivative. It is assumed that $\omega_c \neq 0$. These equations are in the coupled-mode form. The coupling disappears for a constant magnetic field, $\dot{\omega}_c = 0$.

The solutions of (10) become somewhat more transparent if use is

made of the solutions of (29) for the quantities $\mathbf{R}_b$ and $\mathbf{R}_g$, which are given in Ref. 6 and are included in the Appendix:

$$\begin{aligned} \mathbf{R}_b &= (\omega_{c0}/\omega_c)^{\frac{1}{2}} [\mathbf{R}_{b0} F^* e^{i\varphi} - \mathbf{R}_{g0} XF], \\ \mathbf{R}_g &= (\omega_{c0}/\omega_c)^{\frac{1}{2}} [-\mathbf{R}_{b0} (XF)^* e^{i\varphi} + \mathbf{R}_{g0} F]. \end{aligned} \tag{11}$$

For a static field the quantities X, F and φ are functions of z which are dependent only on the spatial variation of ω_c between z_0 and z and the drift velocity of the discs. Their values are determined from (31), (32) and (33). For a spatially varying field, (31) takes the form

$$\frac{dX}{dz} - \frac{i\omega_c(z)X}{v(z)} + \tfrac{1}{2}(1 - X^2) \frac{d \ln [\omega_c(z)/\omega_{c0}]}{dz} = 0 \tag{12}$$

with $X(z_0) = 0$. The quantities F and φ are obtained from the integrals

$$F = \exp\left(-\frac{1}{2}\int_{z_0}^{z} \left\{X(z') \frac{d \ln [\omega_c(z')/\omega_{c0}]}{dz'}\right\} dz'\right) \tag{13}$$

and

$$\varphi = \int_{z_0}^{z} \left[\frac{\omega_c(z')}{v(z')}\right] dz'. \tag{14}$$

In (12), (13) and (14), v can be considered to be independent of z for the small-signal case in the absence of static longitudinal electric fields. If we choose $\omega_m = -\omega_s$, then we may rewrite (11) as

$$\begin{aligned} |\, \omega_s\omega_c \,|^{\frac{1}{2}} \mathbf{R}_b &= |\, \omega_s\omega_{c0} \,|^{\frac{1}{2}} [F^* e^{i\varphi} R_{b0} e^{i\omega_s t_0} - XFR_{g0} e^{i\omega_s t_0}], \\ |\, \omega_s\omega_c \,|^{\frac{1}{2}} \mathbf{R}_g &= |\, \omega_s\omega_{c0} \,|^{\frac{1}{2}} [-(XF)^* e^{i\varphi} R_{b0} e^{i\omega_s t_0} + FR_{g0} e^{i\omega_s t_0}], \end{aligned} \tag{15}$$

which can finally be written in the form

$$\begin{vmatrix} a_c(\omega_s) \\ a_s(-\omega_s) \end{vmatrix} = \begin{vmatrix} F^* e^{i\varphi} & -XF \\ -(XF)^* e^{i\varphi} & F \end{vmatrix} \begin{vmatrix} a_{c0}(\omega_s) \\ a_{s0}(-\omega_s) \end{vmatrix}. \tag{16}$$

We see that, if $\omega_s > 0$, the fast cyclotron wave is coupled to the slow synchronous wave. The determinant of the transformation has the value $|\, F \,|^2(1 - |\, X \,|^2)e^{i\varphi}$, which, in view of (34), has the value $e^{i\varphi}$. Equation (35) shows that the total kinetic power, $|\, a_c \,|^2 - |\, a_s \,|^2 = |\, a_{c0} \,|^2 - |\, a_{s0} \,|^2$, of the two waves is conserved. The quantity $|\, F \,|^2$ represents the kinetic power gain of each wave and $|\, XF \,|^2$ represents the kinetic power gain of one of the waves as a result of an initial excitation of the other. Equation (34) expresses the conservation of kinetic power and the fact that $|\, F \,|^2 \geqq 1$ and $|\, X \,|^2 < 1$. Thus, one may conclude that variations of any kind will always amplify existing cyclotron or synchronous

waves. If both waves exist coherently, one fast and the other slow, then field variations may deamplify the waves. Inspection of (12) shows that, when ω_c is a slow function of z, then X will remain very small and F must remain very close to unity. Hence, for slow changes in the magnetic field strength the kinetic power of each wave is conserved.

It follows that the noise temperature of the cyclotron and synchronous waves is invariant under slow changes in the magnetic field strength. Solutions for X and F for monotonically changing fields are given in Ref. 6. Under the appropriate conditions, X returns periodically to zero. Hence, the noise temperature of the waves may be preserved even for rapid changes in the field if the contour is designed properly.

Sinusoidal field variations of the form $\omega = \omega_{c0}[1 + \Delta \sin 2\pi z/L]$ lead to exponentially growing values of F if $2\pi v/L = \omega_{c0}$. Thus, large amplification is also possible.

IV. AMPLIFICATION IN TIME-VARYING FIELDS

In time-varying fields the quantities X and F will depend upon the initial or entry phase of the disc. An example will be given to demonstrate the phase dependence. It is assumed that the varying component of the magnetic field is a traveling wave, so that one may write

$$\omega_c = \omega_{c0}\{1 + \Delta \sin[\omega_p t - \beta_p(z - z_0)]\}, \tag{17}$$

in which ω_p is the pump frequency and β_p the pump propagation constant. The field near the axis of a cylindrical TE_{01} mode structure is appropriate for this case. It is assumed that $\Delta \ll 1$, so that one may rewrite (31) as

$$\dot{X} - i\omega_{c0}X + \tfrac{1}{2}\Delta(\omega_p - \beta_p v)\cos[\omega_p t - \beta_p(z - z_0)] = 0, \tag{18}$$

in which terms in Δ^2 have been neglected. This corresponds to very weak coupling. Equation (16) has the solution, using the boundary condition $X(t_0) = 0$ and $z - z_0 = v(t - t_0)$,

$$X(\tau) = -\tfrac{1}{2}\Delta(\omega_p - \beta_p v)e^{i\omega_{c0}\tau}\int_0^\tau \cos[\omega_p(\tau' + t_0) - \beta_p v\tau']e^{-i\omega_{c0}\tau'}\,d\tau', \tag{19}$$

in which $\tau = t - t_0$ is the elapsed time in the pump field. Performing the integration, one obtains

$$X(\tau) = -\tfrac{1}{4}\Delta(\omega_p - \beta_p v)\left[\frac{e^{i(\omega_p - \beta_p v - \omega_{c0})\tau} - 1}{i(\omega_p - \beta_p v - \omega_{c0})}\,e^{i\omega_p t_0} - \frac{e^{-i(\omega_p - \beta_p v + \omega_{c0})\tau} - 1}{i(\omega_p - \beta_p v + \omega_{c0})}\,e^{-i\omega_p t_0}\right]e^{i\omega_{c0}\tau}. \tag{20}$$

In the limiting case $\omega_p - \beta_y v = \omega_{c0}$, $\omega_{c0}\tau \gg 1$, one can approximate (20) by

$$X \cong -\tfrac{1}{4}\Delta\omega_{c0}\tau e^{i\omega_{c0}\tau} e^{i\omega_p t_0}. \tag{21}$$

It can be seen that X is a function of the time the electron enters the pump field and increases linearly with drift time. In this limit it can be shown that

$$F \cong e^{\frac{1}{2}(\Delta\omega_{c0}\tau/4)^2}, \tag{22}$$

which satisfies (34), and that $\varphi \cong \omega_{c0}\tau$,

In this case, choosing $\omega_m = \omega_p - \omega_s$, (11) takes the form

$$\begin{aligned} |\omega_s\omega_c|^{\frac{1}{2}}\mathbf{R}_b &= F^* e^{i\varphi} |\omega_s\omega_{c0}|^{\frac{1}{2}} R_{b0} e^{i\omega_s t_0} \\ &\quad - |\omega_s/\omega_m|^{\frac{1}{2}} XF, R_{g0} e^{-i\omega_m t_0} |\omega_m\omega_{c0}|^{\frac{1}{2}}, \\ |\omega_m\omega_c|^{\frac{1}{2}}\mathbf{R}_g &= -|\omega_m/\omega_s|^{\frac{1}{2}} (XF)^* e^{i\varphi} |\omega_s\omega_{c0}|^{\frac{1}{2}} R_{b0} e^{i\omega_s t_0} \\ &\quad + F |\omega_m\omega_{c0}|^{\frac{1}{2}} R_{g0} e^{-i\omega_m t_0}. \end{aligned} \tag{23}$$

Notice that the combination $Xe^{-i\omega_m t_0}$ has a dependence on t_0 of the form $e^{i\omega_s t_0}$ and $X^* e^{i\omega_s t_0}$ has a dependence of the form $e^{-i\omega_m t_0}$. Consequently, one may write

$$\begin{vmatrix} a_c(\omega_s) \\ a_s(\omega_m) \end{vmatrix} = \begin{vmatrix} F^* e^{i\varphi} & -|\omega_s/\omega_m|^{\frac{1}{2}} XF \\ -|\omega_m/\omega_s|^{\frac{1}{2}} (XF)^* e^{i\varphi} & F \end{vmatrix} \cdot \begin{vmatrix} a_{c0}(\omega_s) \\ a_{s0}(\omega_m) \end{vmatrix}, \tag{24}$$

and one may note that the cylotron wave of frequency ω_s is coupled to the synchronous wave of frequency $\omega_m = \omega_p - \omega_s$. If $\omega_p > \omega_s > 0$ then $\omega_m > 0$ and both waves may carry positive kinetic power. Both may also carry negative power when $\omega_p < \omega_s < 0$. If $|\omega_p| < |\omega_s|$ one wave must carry negative kinetic power. Notice first that the transformation represented by (24) reduces to the transformation for the space-varying case given in (16) in the limit $\omega_p = 0$, $\omega_m = -\omega_s$. The transformation still has a determinant of magnitude unity because of the condition given in (34). The significance of this fact can be understood from the following considerations. One may write for the kinetic power output at the signal frequency for unit power input, $P_{ks} = |F|^2$ and the power output at the idler frequency as $P_{km} = |\omega_m/\omega_s|\,|XF|^2$. Note that the Manley-Rowe relationship[10] for a three-frequency reactive enegry converter

$$(P_{ks} - 1) - |\omega_s/\omega_m| P_{km} = 0 \tag{25}$$

is satisfied as a result of the relationship given in (34).

Equation (34) has been seen to play a major role in both stationary and time-varying field pumping. As a result of this condition the kinetic power is conserved in the coupling of a fast (slow) cyclotron wave and a slow (fast) synchronous wave by a spatially varying static magnetic field. In the case of high-frequency pumping, (34) is equivalent to the Manley-Rowe relationship. In fact, conservation of kinetic power and the Manley-Rowe relationship are the same in the special case $\omega_p = 0$. In Ref. 6 it is shown that the condition expressed by (34) is merely a statement of the fact that the momentum canonically conjugate with the angular coordinate is time independent. This implies that the conservation of angular momentum of the total system, beam plus field, and the Manley-Rowe condition are analogous. In line with this fact one may note that the constraints which were placed on the solution, $\omega_p - \beta_p v = \omega_{c0}$ and $\omega_m = \omega_p - \omega_s$, in combination with the unperturbed propagation constants of the cyclotron and synchronous waves, $\beta_s = (\omega_s - \omega_{c0})/v$ and $\beta_m = \omega_m/v$, yield the more commonly known constraints

$$\begin{aligned} \omega_s + \omega_m &= \omega_p , \\ \beta_s + \beta_m &= \beta_p . \end{aligned} \tag{26}$$

It is interesting to note that there are only two coupled modes in the space-varying ($\omega_p = 0$) pumping independent of the degree of coupling, whereas in the high-frequency case there are two modes only in the limit of weak coupling. As the coupling is made stronger, inspection shows that sidebands of mixing between multiples of the pump frequency and the signal frequency are introduced. The idler waves will be both synchronous and cyclotron waves, but the most important waves will be those discussed there. Clearly the difference arises because there can be no multiples of the pump frequency in $\omega_p = 0$ pumping.

V. INFLUENCE OF SPACE CHARGE

In the presence of space charge the motion of a beam electron in the center-of-mass system of a uniform beam can be written approximately as

$$\ddot{\rho} - i(\omega_c\dot{\rho} + \tfrac{1}{2}\dot{\omega}_c\rho) - \tfrac{1}{2}\omega_q^2\rho = 0, \tag{27}$$

in which ω_q is the plasma frequency of the beam. Compare this to (34) for the beam disc. In the limit $\dot{\omega}_c = 0$ this leads to two natural motions of the electron, one at a frequency $\omega_c' = \tfrac{1}{2}\omega_c[1 - (1 - 2\omega_q^2/\omega_c^2)^{\frac{1}{2}}]$ and the other at $\omega_c'' = \tfrac{1}{2}\omega_c[1 + (1 - 2\omega_q^2/\omega_c^2)^{\frac{1}{2}}]$. In the limit of zero space charge, $\omega_q = 0$, these motions correspond to the guiding center and orbital motion.

In order to amplify the internal motion of the electrons it is necessary to pump at either ω_c' or ω_c'' rather than ω_c. Hence, if $2\omega_q^2/\omega_c^2$ is sufficiently close to one, then both ω_c' and ω_c'' are appreciably different from ω_c and pumping at a frequency equal to ω_c will amplify the beam motion and consequently the beam waves without amplifying the electron orbits as has been shown by Adler et al. for quadrupole pump fields.[11]

VI. CONCLUSION

The cyclotron wave associated with the rotational motion of the beam electrons and the synchronous wave associated with the spatial configuration of the beam have been shown to be coupled by varying magnetic fields. Spatial variations of any kind in the strength of the magnetic field always amplify existing cyclotron or synchronous waves. One of the coupled modes is fast and the other slow, and the total kinetic power is conserved. Sinusoidal variations in the field strength such that $2\pi v/L = \omega_{c0}$, in which v is the drift velocity and L is the periodicity, are particularly effective in coupling the modes. The kinetic power in each mode is conserved when the field varies slowly.

In magnetic pumping with a time-varying field, a fast cyclotron wave and a fast synchronous wave are coupled if the signal frequency is lower than the pump frequency. If the signal frequency is larger than the pump frequency one of the modes must be slow. The Manley-Rowe condition is satisfied by this interaction. The space-varying magnetic field is seen to be a special case of parametric pumping at zero pump frequency. Conservation of kinetic power in this case is a special case of the Manley-Rowe relationship.

VII. ACKNOWLEDGEMENTS

The author would like to acknowledge the interest and stimulation provided by J. Feinstein, S. J. Buchsbaum and J. W. Klüver.

APPENDIX

The azimuthally symmetric fields are derived from a vector potential, $\mathbf{A}$, with rectangular components $(-\frac{1}{2}yH(z,t), \frac{1}{2}xH(z,t), 0)$ which give rise to an electric field, $\mathbf{E} = -c^{-1}\partial\mathbf{A}/\partial t$, with components $(\frac{1}{2}c^{-1}y\partial H/\partial t, -\frac{1}{2}c^{-1}x\partial H/\partial t, 0)$ and a magnetic field, $\mathbf{H} = \text{curl } \mathbf{A}$, with components $(-\frac{1}{2}x\partial H/\partial z, -\frac{1}{2}y\partial H/\partial z, H)$. The equation of motion for the beam disc is

$$\begin{aligned} m\ddot{x} &= -(e/c)[\tfrac{1}{2}y(\partial H/\partial t + \dot{z}\partial H/\partial z) + \dot{y}H], \\ m\ddot{y} &= (e/c)[\tfrac{1}{2}x(\partial H/\partial t + \dot{z}\partial H/\partial z) + \dot{x}H]. \end{aligned} \tag{28}$$

Multiplying the second equation by i and writing $\mathbf{R} = x + iy$ yields

$$\ddot{\mathbf{R}} - i(\omega_c\dot{\mathbf{R}} + \tfrac{1}{2}\dot{\omega}_c\mathbf{R}) = 0, \tag{29}$$

in which $\omega_c(z,t) = eH(z,t)/mc$. Assuming that the time dependence of the z-coordinate of the disc can be specified the solution for (29) may be written

$$\begin{aligned}\mathbf{R}_b(t) &= \dot{\mathbf{R}}(t)/i\omega_c = (\omega_{c0}/\omega_c)^{\frac{1}{2}}[\mathbf{R}_{b0}F^*e^{i\varphi} - \mathbf{R}_{g0}XF],\\ \mathbf{R}_g(t) &= \mathbf{R}(t) - \mathbf{R}_b(t) = (\omega_{c0}/\omega_c)^{\frac{1}{2}}[-\mathbf{R}_{b0}(XF)^*e^{i\varphi} + \mathbf{R}_{g0}F],\end{aligned} \tag{30}$$

in which X satisfies the first-order differential equation

$$\dot{X} - i\omega_cX + \tfrac{1}{2}\dot{\omega}_c(1 - X^2)/\omega_c = 0 \tag{31}$$

with initial conditions $X(t_0) = 0$, $\mathbf{R}_{b0} = \mathbf{R}_b(t_0)$, $\mathbf{R}_{g0} = \mathbf{R}_g(t_0)$,

$$\varphi(t) = \int_{t_0}^{t} \omega_c(t')\,dt' \tag{32}$$

and

$$F(t) = \exp\left\{-\frac{1}{2}\int_{t_0}^{t}[\dot{\omega}_cX/\omega_c]\,dt'\right\}. \tag{33}$$

Multiplying (31) by X^* and the complex conjugate equation by X, adding and integrating yields

$$|F|^2(1 - |X|^2) = 1. \tag{34}$$

Combining (30) and (34) yields

$$\omega_c(R_g^2 - R_b^2) = \omega_{c0}(R_{g0}^2 - R_{b0}^2). \tag{35}$$

REFERENCES

1. Adler, R., Parametric Amplification of the Fast Electron Wave, Proc. I.R.E., **46**, 1958, p. 1300.
2. Adler, R., Hrbek, G. and Wade, G., A Low-Noise Electron-Beam Parametric Amplifier, Proc. I.R.E., **46**, 1958, p. 1756.
3. Adler, R., Hrbek, G. and Wade, G., The Quadrupole Amplifier, a Low-Noise Parametric Device, Proc. I.R.E., **47**, 1959, p. 1713.
4. Gordon, E. I., Buchsbaum, S. J. and Feinstein, J., A Transverse Field Amplifier Employing Cyclotron Resonance Interaction, 17th Conf. on Electron Tube Research, Mexico City, June 1959.
5. Gordon, E. I., A Transverse Field Traveling Wave Tube, Proc. I.R.E., **48**, 1960, p. 1158.
6. Gordon, E. I., Charged Particle Orbits in Varying Magnetic Fields, J. Appl. Phys., **31**, 1960, p. 1187.
7. Bobroff, D. L., Independent Space Variables for Small Signal Electron Beam Analysis, I.R.E. Trans., **ED-6**, 1959, p. 68.
8. Siegman, A. E., Waves on a Filamentary Electron Beam in a Transverse Field Slow-Wave Circuit, J. Appl. Phys., **31**, 1960, p. 17.
9. Klüver, J. W., to be published.
10. Manley, S. M. and Rowe, H. E., Some General Properties of Nonlinear Elements, Proc. I.R.E., **44**, 1956, p. 904.
11. Adler, R., Ashkin, A. and Gordon, E. I., 18th Conf. on Electron Tube Research, Seattle, Wash., June 1960.

"Ionic Radii," Spin-Orbit Coupling and the Geometrical Stability of Inorganic Complexes*

By ANDREW D. LIEHR

(Manuscript received December 24, 1959)

The van Santen and van Wieringen theory of "ionic radii" is briefly reviewed and extended to include spin-orbit influences. It is noted that, although for the most part spin-orbit forces have little effect upon stereochemical predictions made for the first transition series, noteworthy exceptions to this rule occur for octahedral complexes of Co^{++} and tetrahedral complexes of Ni^{++} and Cu^{++}. Indeed, it is found that spin-orbit coercions render the ground states of these molecules Jahn-Teller resistant. Diagrams are displayed and tables compiled to illustrate the variation of ionic radii with atomic number for the second and third transition series. Paths for future theoretical research are indicated and an exhortation for closer theoretical-experimental alliance is promulged.

Now, although the ligand field theory is in its 30th year, it has only been within the last decade that its chemical fruits have been earnestly harvested. One of the earliest pickings of its ripened orchards was accomplished by van Santen and van Wieringen.[1] These researchers noticed that the enigmatic irregularities of the existent transition metal "ionic-radii" tabulations could be neatly rationalized on the basis of the Bethe-Kramers-Van Vleck[2,3,4] crystalline field formalism. I now wish to extend their argument to include metals which possess a nonnegligible spin-orbit interaction and to point out certain modifications (due to Jahn[5]) of the Jahn-Teller stability rules[6] which such interactions engender. But before we may embark upon this tour of the crystallographic implications of the existence of spin-orbit forces in inorganic complexes, it is necessary to review briefly the basic precepts of ligand (or crystalline) field theory.

* Presented at the 137th meeting of the American Chemical Society, Symposium on the Synthesis and Properties of Inorganic Compounds, Cleveland, April 5–14, 1960.

If a positive rare gas core is placed in the midst of an octahedral arrangement of negative (or neutral) ligands (see Fig. 1), a mutual electrical polarization of these two sets of charge clouds occurs. With the concomitant shifting of ligand electrical charge toward the central metal ion, driven by the "charge neutrality" forces, the ligand molecular orbitals attain a partially metallic character. Conversely, the central metal ion orbitals, *both occupied and unoccupied*, partake of the surrounding ligand charge density distribution (Fig. 1).* *This reciprocity of electronic charge induces a separation of the as yet unoccupied nd-like, (n = 3, 4, 5), orbitals of the central (rare gas-like) ion.* In covalent bonding terminology these latter orbitals become antibonding in nature, dividing themselves into two groups: the σ-antibonding e_g molecular orbitals, which originate from the $nd_{x^2-y^2}$ and nd_{z^2} atomic electron distributions, and the π-antibonding t_{2g} molecular orbitals, which proceed from the atomic nd_{xy}, nd_{yz}, and nd_{xz} charge dispositions (see Fig. 2).

When the atomic number of the metallic core is increased so that additional electrons become available to the complex, the empty antibonding t_{2g} and e_g orbitals are sequentially filled. Now, from general electro-

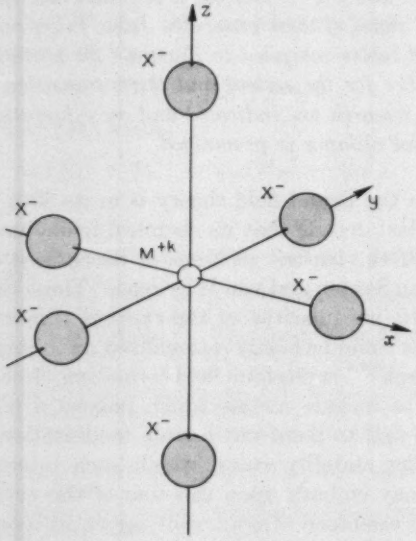


Fig. 1 — Geometrical disposition of an octahedral complex.

* We should here like to suggest the use of the words *addend* (that which adds) and *augend* (that which augments) as appropriate synonyms for ligand (that which ligates).

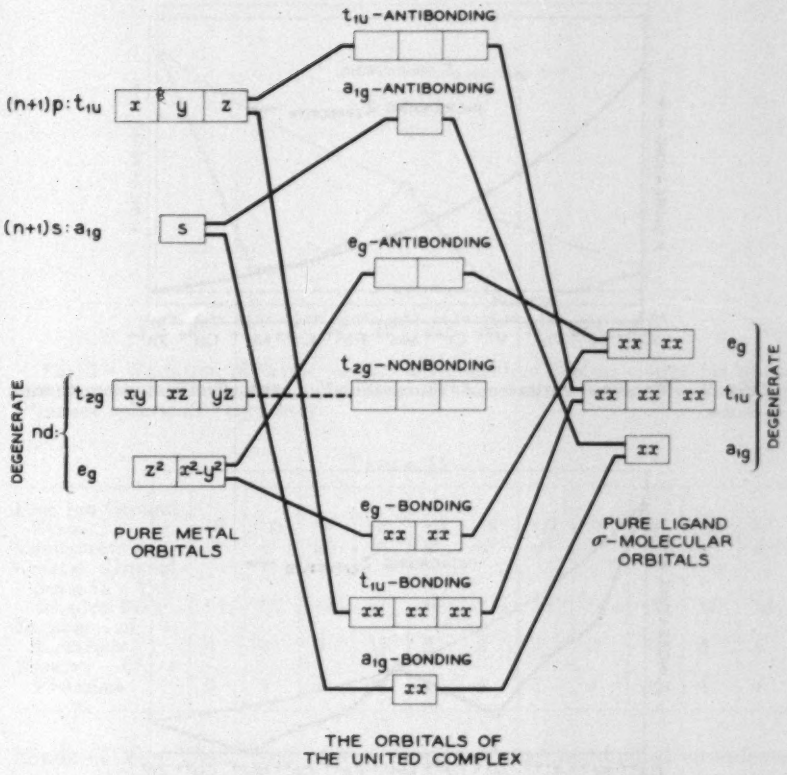


Fig. 2 — Energy levels of a σ-bonded octahedral complex. When π-bonding is added, the empty t_{2g} orbital becomes π-antibonding in character.

static considerations we would expect an attendant smooth (monotonic) decrease of "ionic radii" with the increasing charge of the core, as is shown in Fig. 3. But this is not the case in actuality! "Real" ionic radii vary in the jagged fashion pictured in Fig. 4. What van Santen and van Wieringen were able to show is that the observed irregularity of the "ionic radii" behavior was due to the serial occupancy of the antibonding t_{2g} and e_g orbitals: that the occupancy of the weakly antibonding *xy*, *xz*, and *yz*-type orbitals allowed a *contraction* of the bonding radii and that the population of the strongly antibonding x^2- y^2- and z^2-type orbitals forced an *expansion* of the bonding radii for octahedral complexes. This situation is emphasized by the registration of their results given in Table I. In addition, the theorem of Jahn and Teller, in the

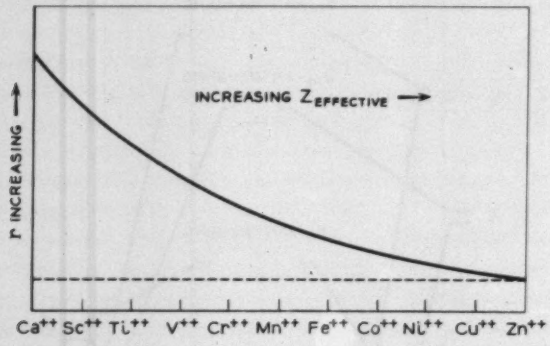


Fig. 3 — Expected variation of "ionic radius", r, with effective nuclear charge, $Z_{effective}$.

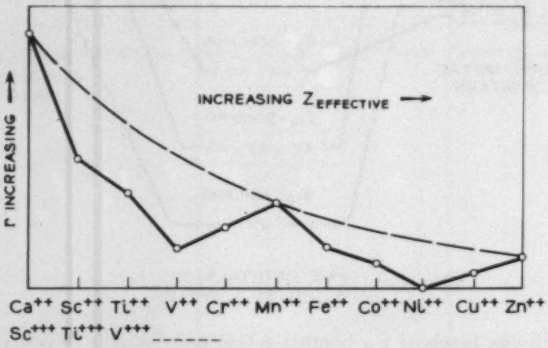


Fig. 4 — Variation of "ionic radius" with effective nuclear charge for octahedral geometries. The maxima and minima of the curves are states of perfect cubic ("pseudo spherical") symmetry.

TABLE I

Free Ion Ground State	^1S	^2D	^3F	^4F	^5D	^6S	^5D	^4F	^3F	^2D	^1S
Configuration	d^0	d^1	d^2	d^3	d^4	d^5	d^6	d^7	d^8	d^9	d^{10}
Crystal Ground State in an Octahedral Field	$^1A_{1g}$	$^2T_{2g}$	$^3T_{1g}$	$^4A_{2g}$	^5E_g	$^6A_{1g}$	$^5T_{2g}$	$^4T_{1g}$	$^3A_{2g}$	^2E_g	$^1A_{1g}$
Number of e_g Electrons	0	0	$\frac{1}{5}$	0	1	2	2	$2\frac{1}{5}$	2	3	4
Number of t_{2g} Electrons	0	1	$1\frac{4}{5}$	3	3	3	4	$4\frac{4}{5}$	6	6	6

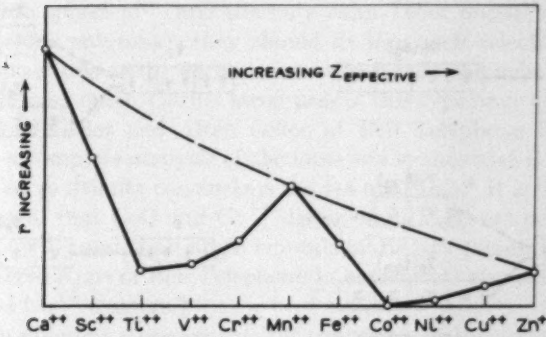


Fig. 5 — Variation of "ionic radius" with effective nuclear charge for tetrahedral environs. The maxima and minima of the curves are states of perfect cubic ("pseudo spherical") eurythmy.

TABLE II

Free Ion Ground State	^1S	^2D	^3F	^4F	^5D	^6S	^5D	^4F	^3F	^2D	^1S
Configuration	d^0	d^1	d^2	d^3	d^4	d^5	d^6	d^7	d^8	d^9	d^{10}
Crystal Ground State in a Tetrahedral Field	$^1A_1$	^2E	$^3A_2$	$^4T_1$	$^5T_2$	$^6A_1$	^5E	$^4A_2$	$^3T_1$	$^2T_2$	$^1A_1$
Number of t_2 Electrons	0	0	0	$1\frac{1}{3}$	2	3	3	3	$4\frac{1}{3}$	5	6
Number of e Electrons	0	1	2	$1\frac{2}{3}$	2	2	3	4	$3\frac{2}{3}$	4	4

hands of Van Vleck[7] and others, predicted that octahedral complexes with an odd number of electrons in the antibonding e_g shell [e.g. Cr^{++} and $Mn^{+++}(d^4)$ and $Cu^{++}(d^9)$] should be permanently distorted. Conclusions of a similar nature were also obtained for tetrahedral complexes (Fig. 5 and Table II).

How does the complication of spin-orbit coupling modify these results? Simply in this way: The linkage of the spin and orbital magnetic moments serves to split the six-fold degenerate t_{2g} states [three possible orbital motions and two possible spin directions] into a four-fold level called $\gamma_8(t_{2g})$ (by Bethe[2]) and a two-fold level called $\gamma_7(t_{2g})$. The four-fold multiple of spin-orbit states, e_g, remains intact, and also takes on the new label $\gamma_8(e_g)$. The sequence of antibonding orbitals in octahedral and tetrahedral complexes then becomes as shown in Fig. 6.*

* At this juncture, we should like to propose that the Mulliken notation, (a, e, t, etc.), be reserved for pure *orbital* states, and the Bethe notation, γ_j, ($j = 1, 2, \ldots, 8$), for *spin-orbital* configurations.

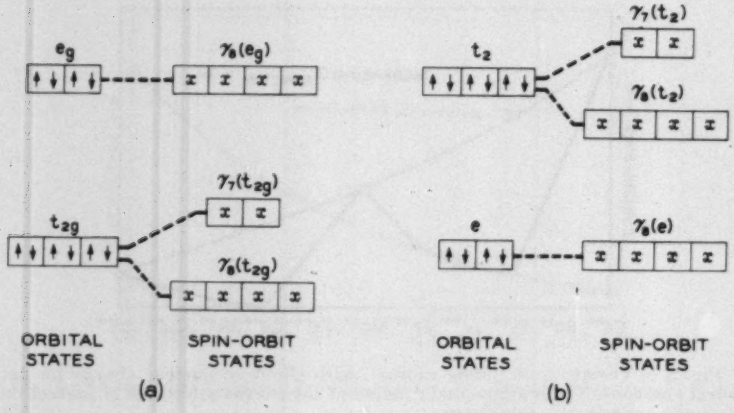


Fig. 6 — Correlation of the orbital and spin-orbit states of (a) octahedral and (b) tetrahedral inorganic complexes.

For the most part, the entrance of spin-orbit correlations has no effect upon stereochemical predictions made for the first transition series. Notable exceptions to this rule occur for octahedral complexes of Co^{++} and tetrahedral complexes of Ni^{++} and Cu^{++}. Both octahedral Co^{++}, which subsumes the configuration $\gamma_8(t_{2g})^4\gamma_7(t_{2g})^1\gamma_8(e_g)^2$ in our naïve theory, and tetrahedral Cu^{++}, which takes on the assignment

$$\gamma_8(e_g)^4\gamma_8(t_{2g})^4\gamma_7(t_{2g})^1,$$

exist in the Jahn-Teller *resistant* electronic dispositions, Γ_6 and Γ_7 respectively [according to Jahn,[5] the doubly degenerate (cubic) spin-orbit states of Bethe, Γ_6 and Γ_7, the so-called Kramers doublets, can *not* exhibit (pure) Jahn-Teller conformational instability]. Since the nearest-lying electronic states are ~400 (Shulman[8]) and ~1000 cm^{-1} (Liehr[9]) away for Co^{++} and Cu^{++}, respectively, their ground electronic states are also *not* susceptible to large pseudo Jahn-Teller coercions, and should thus be stable in the regular polyhedral arrangement. Similarly, tetrahedral Ni^{++}, which has the electron distribution $\gamma_8(e_g)^4\gamma_8(t_{2g})^4$ in the simple one-electron scheme, exists in the totally symmetric (nondegenerate) state, Γ_1, in which the Jahn-Teller forces are again inoperative. Since the closest electronic disposition, which under nuclear displacements might perturb the ground state conformational regularity, is ~300 cm^{-1} distant (Liehr and Ballhausen[10]), Ni^{++} should also *not* exhibit any large pseudo Jahn-Teller distortions, and thus should be stable in the regular tetrahedral form. Moreover, since the Bethe states Γ_1, Γ_6 and Γ_7

are "pseudo spherical" (and the only Jahn-Teller impotent states possible for cubic environs), they should fit into such selective crystallographic structures as the garnets, into which only the most highly symmetric ions may enter. Garnet structures of this type have been prepared by Seymour Geller and Alten Gilleo at Bell Telephone Laboratories; however, a complete analysis of the ionic site symmetries is still in progress, and so no definite conclusions are yet available.* It is interesting to note, though, that CoO and Co^{++} dissolved in MgO are perfectly octahedral (Low[11]), and that single crystals of $KCoF_3$ (recently characterized by Kerro Knox of Bell Telephone Laboratories) also contain regular octahedral Co^{++} clusters.†

A much different story unfolds for complexes of the second and third transition series. Here the separation of the $\gamma_7(t_{2g})$ and $\gamma_8(t_{2g})$ levels which arise from the π-antibonding (orbital) t_{2g} state of octahedral complexes may become large enough so that electronic repulsions do not disturb the orderly filing of the levels. Hence, the orbitals tend to fill up with the first four electrons going into the lowest $\gamma_8(t_{2g})$ level, the next two into the $\gamma_7(t_{2g})$ orbit, and the next four into the highest $\gamma_8(e_g)$ level, for octahedral complexes; and the two $\gamma_8(e)$ and $\gamma_8(t_2)$ trajectories are occupied first for tetrahedral complexes and the $\gamma_7(t_2)$ level last. In Figs. 7 and 8 and Tables III and IV we summarize the structural chem-

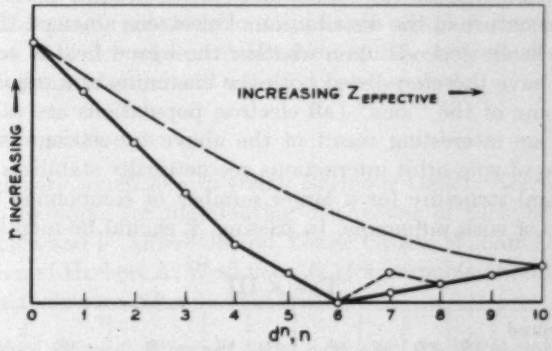


Fig. 7 — Variation of "ionic radius" with effective nuclear charge for medium-strong spin-orbit coupled octahedral complexes. The configurations d^n, (n = 0,4,5,6,7,10) are states of perfect cubic ("pseudo spherical") symmetry. The states d^3 and d^8 are but slightly asymmetric due to their spin distributions.

* Co^{++} is now known to enter into octahedral garnet sites (Geller[12]).

† The large reduction in magnitude of the Jahn-Teller forces in FeO and Fe^{++} dissolved in MgO (Low[11]), and in $KFeF_3$ (Knox[13]) is also explicable on the basis of a (partial) spin-orbit stabilization (Van Vleck[14]).

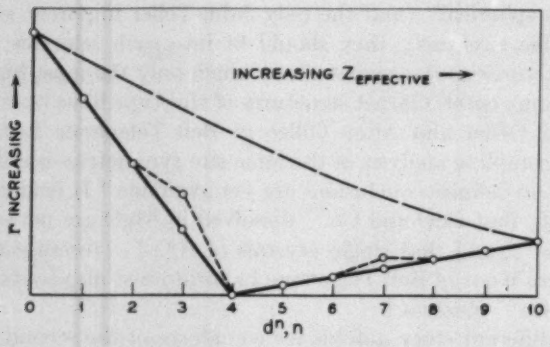


Fig. 8 — Variation of "ionic radius" with effective nuclear charge for medium-strong spin-orbit coupled tetrahedral complexes. The configurations d^n, (n = 0,4,8,9,10) are states of perfect cubic ("pseudo spherical") eurythmy. The states d^2 and d^7 are but slightly asymmetric due to their spin distributions [the same may be said for d^5 if it exhibits *maximum* spin].

ical expectations for ions of the second and third (especially the third) transition series. The ups and downs of the radii plots in these figures follow immediately upon the determination of the population of the t_{2g}-like and e_g-like spin-orbitals $\gamma_8(t_{2g})$, $\gamma_7(t_{2g})$, and $\gamma_8(e_g)$, just as was the case for the original van Santen and van Wieringen theory.

The exact nature of the distribution of electrons amongst the various tetrahedral levels depends upon whether the ligand field is very strong or not. We have therefore listed both the maximum and minimum spin cases for some of the "ions" (all electron populations are taken to be integral). One interesting result of the above tabulations is that the introduction of spin-orbit interactions geometrically stabilizes the regular polyhedral structure for a larger number of compounds than does the absence of such influences. In passing, it should be mentioned that

TABLE III

Free Ion Ground State	^1S	^2D	^3F	^4F	^5D	^6S	^5D	^4F	^3F	^2D	^1S
Configuration	d^0	d^1	d^2	d^3	d^4	d^5	d^6	d^7	d^8	d^9	d^{10}
Crystal Ground State in an Octahedral Field	Γ_1	Γ_8	Γ_5	Γ_8	Γ_1	Γ_7	Γ_1	Γ_5 or Γ_3	Γ_3	Γ_8	Γ_1
Number of $\gamma_8(e_g)$ Electrons	0	0	0	0	0	0	0	2 or 1	2	3	4
Number of $\gamma_7(t_{2g})$ Electrons	0	0	0	0	0	1	2	1 or 2	2	2	2
Number of $\gamma_8(t_{2g})$ Electrons	0	1	2	3	4	4	4	4	4	4	4

TABLE IV

Free Ion Ground State	^1S	^2D	^3F	^4F	^5D	^6S	^5D	^4F	^3F	^2D	^1S
Configuration	d^0	d^1	d^2	d^3	d^4	d^5	d^6	d^7	d^8	d^9	d^{10}
Crystal Ground State in a Tetrahedral Field	Γ_1	Γ_3	Γ_5	Γ_3	Γ_1	Γ_3	Γ_3 or Γ_5 (?)	Γ_3	Γ_1	Γ_7	Γ_1
Number of $\gamma_7(t_2)$ Electrons	0	0	0	0	0	0	0	1 or 0	0	1	2
Number of $\gamma_8(t_2)$ Electrons	0	0	0	1 or 0	0	1	2	2 or 3	4	4	4
Number of $\gamma_3(e)$ Electrons	0	1	2	2 or 3	4	4	4	4	4	4	4

a possible test of the foretold octahedral "ionic radii" demeanor would be afforded by the structural determination of the gaseous WF_6, ReF_6, etc., series of compounds.

To firmly pin down the stereochemical deportment of transition metal complexes, a good deal more work must be done, especially upon the d^n, $(n = 3, 4, 6, 7)$, configurations for strong spin-orbit forces. Carl Ballhausen and myself are presently investigating the $d^{3,7}$ case; we hope that others will extend our results to encompass the $d^{4,6}$ situation also. At this point we are in a position to witness a veritable mushrooming of theoretical and theoretically inspired experimental research in the field of transition metal, rare earth and actinide complex chemistry. I sincerely hope that this close collaboration of the theoretical and experimental inorganic chemist will indeed come to pass.*

ACKNOWLEDGMENTS

I should very much like to thank Seymour Geller, Kerro Knox and Robert G. Shulman for illuminating discussions of their researches on similar topics, and F. Albert Cotton, Dieter Gruen, William Low, Ronald S. Nyholm and Herbert A. Weakliem, Jr., for prepublication information concerning their researches in this area.

* *Note added in proof:* A revised set of ionic radii for M^{+2} ions has been formulated by Geller,[15] and for M^{+2} ions by Knox,[13] which agree quite nicely with the theory outlined in this paper. Mr. Geller informs me that the Ti^{+3} radius reported in his article should, on the basis of recent measurements[16] on $LaTiO_3$, now read 0.633 angstrom. Also, it is interesting to note that Gill and Nyholm[17] have recently prepared some truly tetrahedral complexes of divalent nickel, which serve to substantiate the results obtained in melts of $NiCl_2$.[18,19] The present status of tetrahedral Ni^{++} chemistry is completely outlined in the recent paper of Cotton, Faut and Goodgame.[20] Racah, Schonfeld and Low[21] and Weakliem[22] have recently completed the "exact" calculation of the energy levels of $d^{3,7}$ complexes; therefore, the original computation planned for these systems by Carl Ballhausen and myself has understandably been abandoned.

REFERENCES

1. van Santen, J. H. and van Wieringen, J. S., Rec. Trav. Chim., **71,** 1952, p. 420.
2. Bethe, H., Ann. der Physik, **3,** 1929, p. 133.
3. Kramers, H. A., Proc. Amsterdam Acad., **32,** 1929, p. 1176.
4. Van Vleck, J. H., *The Theory of Electric and Magnetic Susceptibilities*, Clarendon Press, Oxford, 1932, p. 287.
5. Jahn, H. A., Proc. Roy. Soc., **164A,** 1938, p. 117.
6. Jahn, H. A., and Teller, E., Proc. Roy. Soc., **161A,** 1937, p. 220.
7. Van Vleck, J. H., J. Chem. Phys., **7,** 1939, pp. 61; 72.
8. Shulman, R. G., Phys. Rev. Letters, **2,** 1959, p. 459.
9. Liehr, A. D., J. Phys. Chem., **64,** 1960, p. 43.
10. Liehr, A. D. and Ballhausen, C. J., Ann. Physics, **6,** 1959, p. 134.
11. Low, W., Phys. Rev., **109,** 1958, p. 256.
12. Geller, S., J. Appl. Phys., **31,** Supplement, 1960, p. 30S.
13. Knox, K., private communication.
14. Van Vleck, J. H., Disc. Far. Soc., No. 26, 1958, p. 96; Physica, **26,** 1960, p. 544.
15. Geller, S., Acta Cryst., **10,** 1957, p. 248.
16. Johnston, W. D. and Sestrich, D. E., private communication to S. Geller.
17. Gill, N. S. and Nyholm, R. S., J. Chem. Soc., **1959,** p. 3997.
18. Gruen, D. M. and McBeth, R. L., J. Phys. Chem., **63,** 1959, p. 393.
19. Sundheim, B. R. and Harrington, G., J. Chem. Phys., **31,** 1959, p. 700.
20. Cotton, F. A., Faut, O. D. and Goodgame, D. M. L., J. Am. Chem. Soc., in press.
21. Racah, G., Schonfeld, Miss G. and Low, W., private communication; Schonfeld, Miss G., Thesis, Hebrew Univ., Israel, 1959.
22. Weakliem, H. A., Jr., private communication.

ADDITIONAL BIBLIOGRAPHY

23. Dunitz, J. D. and Orgel, L. E., *Advances in Inorganic Chemistry and Radiochemistry*, Academic Press, New York, 1960, Vol. 2, pp. 1–60.
24. Hush, N. S. and Pryce, M. H. L., J. Chem. Phys., **26,** 1957, p. 143.
25. Hush, N. S. and Pryce, M. H. L., J. Chem. Phys., **28,** 1958, p. 244.
26. Hush, N. S., Dis. Far. Soc., No. 26, 1958, p. 145.
27. Liehr, A. D. and Ballhausen, C. J., Ann. Physics, **3,** 1958, p. 304.
28. McClure, D. S., *Solid State Physics*, Academic Press, New York, 1959, Vol. 9, pp. 399–525.
29. Moffitt, W. E., Goodman, G. L., Fred, M. and Weinstock, B., Mol. Phys., **2,** 1959, p. 109.
30. Öpik, U. and Pryce, M. H. L., Proc. Roy. Soc., **238A,** 1957, p. 425.

Molecular Structure in Crystal Aggregates of Linear Polyethylene

By R. D. BURBANK

(Manuscript received July 6, 1960)

Crystal aggregates of linear polyethylene have been studied in the electron microscope. Twinning has been observed to occur across (530) planes, and possibly across (120) planes. Crystal morphologies have been observed which exhibit (530) and (540) faces. Electron interference effects have been observed which give rise to contrast lines and figures which frequently are parallel to crystallographic directions. These observations, and those of others, have been interpreted in terms of recently proposed ideas on molecular chain folding. It is suggested that chain folds may lie in a variety of fold planes or fold surfaces which are normal, or nearly normal, to the crystal lamellae. It is shown that a continuity of fold structure is necessary and possible across a wide variety of boundaries delineating regions of different fold structure. It is shown that these structural concepts are compatible with recently proposed ideas on the growth of lamellar polymer crystals and can suggest new details of the growth process.

I. INTRODUCTION

A variety of properties — chemical, electrical and mechanical — have combined to make polyethylene an extremely important material in the telephone system. The sustained attempt to improve the characteristics of polyethylene for different applications depends very much on further understanding of the fundamental structure and behavior of high polymer molecules. Unfortunately, high polymers manifest a tremendous range of structural organization, almost comparable in complexity with the simpler biological systems. Any simplification that can be obtained in the structural organization of a polymer is of inestimable value if the material is to be used for fundamental study.

The discovery that well-developed crystals of linear polymers can readily be obtained from dilute solutions was made independently in several laboratories.[1,2,3] A lamellar-like habit and spiral growth terraces

were found to be ubiquitous features of the crystals. It was recognized that three-dimensional crystal growth must occur by the Frank screw-dislocation mechanism. In addition, Keller[2] concluded that the molecules in linear polyethylene crystals must sharply fold back on themselves with a chain length of about 100 angstroms between folds. A similar deduction concerning molecular folding in films of gutta percha[4] had long been forgotten until Keller called attention to it. Keller and O'Connor[5] postulated that the molecular folds in linear polyethylene lie in (110) planes in crystals which are bounded by (110) faces. They also obtained electron diffraction patterns showing twinning on (110) planes.

With improved techniques to minimize electron beam damage, Agar, Frank and Keller[6] obtained a profusion of electron interference effects with linear polyethylene crystals. Moiré fringes, Bragg extinction contours, and other contrast effects were obtained in bright field and dark field illumination. The surface morphology of some crystals bounded by (110) faces indicated a distinct division into four quadrants. It was proposed that the molecular folds are arranged in four different sets of (110) planes in the various quadrants. Consistent with this proposal was the observation that adjacent quadrants satisfied different conditions of diffraction. Meanwhile, Sella and Trillat[7] reported electron diffraction patterns showing twinning of polyethylene on both (110) and (310) planes, a variety of multiple diffraction effects and striking contrast periodicities in microfibrils.

Bassett, Frank and Keller[8] and Niegisch[9] independently proposed that lamellae of linear polyethylene grow as hollow, nonplanar pyramids which collapse upon drying to leave planar lamellae with a characteristic surface morphology involving a central triangular pleat or fold. Bassett et al. observed that in crystals bounded by both (110) and (100) faces a structure may be observed in the electron microscope with dark field illumination which divides the crystal into six sectors. It was suggested that, in the sectors bounded by (100) faces, the molecular folds were parallel to (100) planes. This concept was supported by the observation that the (100) sectors have a different melting point or transformation temperature than the rest of the crystal.

Niegisch and Swan[10] and Reneker and Geil[11] observed polyethylene lamellae which have folded over along the short diagonal parallel to the b axis. The fold edge is angular rather than straight. This is consistent with a hollow pyramid interpretation and permits a direct measurement of the inclination of such a pyramid to the basal plane. Both Niegisch and Swan and Reneker and Geil have shown that there is an arrangement of folds in (110) planes which forms a (111) surface. The hollow pyramid

formed by four (111) surfaces has an inclination in agreement with that deduced from observations of folded lamellae. Reneker and Geil[11] have extended the argument considerably, however. They show that there are other fold arrangements in (110) planes which form (001) or (112) surfaces. They have interpreted other lamellar morphologies in which corrugations along (310) planes and (530) planes are prominent.

Recently the writer had access to a portion of the operating time on a Siemens Elmiskop I electron microscope for a period of several months. A number of observations were made on crystal aggregates of linear polyethylene. Certain crystallographic relationships were found that have been interpreted in terms of the recently developed ideas on fold structures. Many speculations concerning fold structures have been made in a systematic way that should be useful in discussing molecular configurations of crystalline polymers in general.

II. EXPERIMENTAL

The linear polyethylene used throughout was Marlex 50, melt index 6.0. Crystals were grown from xylene solutions at 0.04 and 0.07 per cent weight concentrations. Cooling rates were not carefully controlled. The heat under a small oil bath was simply removed and the solutions cooled to room temperature over a period of about 60 to 90 minutes. Specimens were prepared by stirring up the crystal-liquid suspension, removing a small quantity of suspension in a fine dropper, and depositing a few small drops on a specimen support. The specimen supports consisted of evaporated carbon films which were mounted on fine copper mesh.

Experimental details given by Agar et al.,[6] were most helpful as a guide to minimizing electron-beam damage in the Siemens Elmiskop I. The double condenser was used throughout the work. In general, one would not expect a reciprocity law to hold between intensity and time. Rather, one would expect low intensities and long times to produce less damage than high intensities and short times. Belavtseva[12] has verified this by experiment, and he has also found higher energy electrons to be less damaging. However, 100-kv electrons never gave as good contrast as 80-kv electrons. Consequently, 80-kv beams were used for most of the work. Best results were obtained with 100μ or 50μ condenser apertures, 50μ or 20μ objective apertures, 2500$\times$ magnification, completely defocused condensers and very long exposure times. Specimens exposed for a 3-minute dark field micrograph, 45-second diffraction pattern and 2-minute bright field micrograph still gave excellent interference effects. Instrumental stability appeared to be adequate at all times. Considerable darkroom adaption was necessary for the low intensities involved.

An interesting diffraction pattern obtained early in the investigation is illustrated in Fig. 1. Discounting some very fine spots which are clustered on (110) and (200) rings, all of the reflections can be accounted for by twinning on (530) and extensive double diffraction. The measured angle between the twinned lattices is 83°55′. Measurement was made with a circular measuring device capable of reading to 2.5′, such as is used for X-ray precession camera photographs. The calculated angle is 83°58′, assuming an a/b axial ratio of 1.5. The axial ratio was measured on all diffraction patterns, and angular relations were not sought unless the measured ratio was 1.5. As the crystals deteriorate in the electron beam the axial ratio increases and values up to 1.65 were observed. Axial ratio measurements could conceivably form an empirical yardstick for quantitative studies of electron beam damage. Fig. 2 is a bright-field micrograph of the (530) twins. Unfortunately, this specimen was destroyed in the beam. The crystals are "dead" and give no interference effects. There is a confused area in the center of the micrograph where both orientations are intergrown.

Fig. 3 is a bright-field micrograph of interpenetrant crystal growth. There are two large branches which form a cross. A smaller branch has

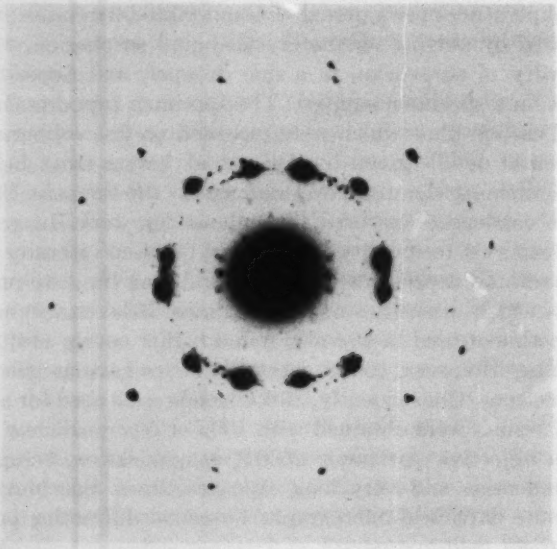


Fig. 1 — Electron diffraction pattern given by crystals twinned across (530) plane.

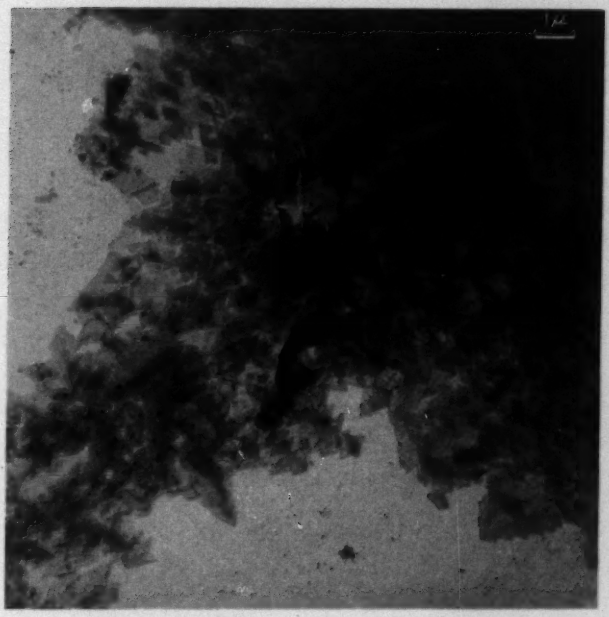


Fig. 2 — Bright-field micrograph of crystals twinned across (530) plane.

grown out from the upper right area of the cross. Fig. 4 is the diffraction pattern of this aggregate. Discounting a scattering of very fine spots, there are two intense arrays of reflections and a third array of moderate intensity. The measured angle between the lattices of intense spots is 68°40′. The calculated angle for (110) twinning is 67°2′. The measured angle between the lattice of moderate intensity and the right-hand lattice of high intensity is 35°22′. The calculated angle for (120) twinning is 36°38′. A variety of weaker reflections caused by double diffraction can be recognized. The crystals illustrated in Figs. 3 and 4 were obtained from 0.07 per cent xylene solution; all the other specimens under discussion were obtained from 0.04 per cent xylene solution.

Fig. 5 is a dark-field micrograph of a type of crystal aggregate which is very common in 0.04 per cent solutions. Many of the specimens illustrated by Agar et al.,[6] are of this type. Similar interference effects may be noted throughout. We should like to direct attention to the Bragg extinction contours which occur in profusion; they may appear as curves, loops, parallelograms, quadrilaterals, or straight lines. The contours of-

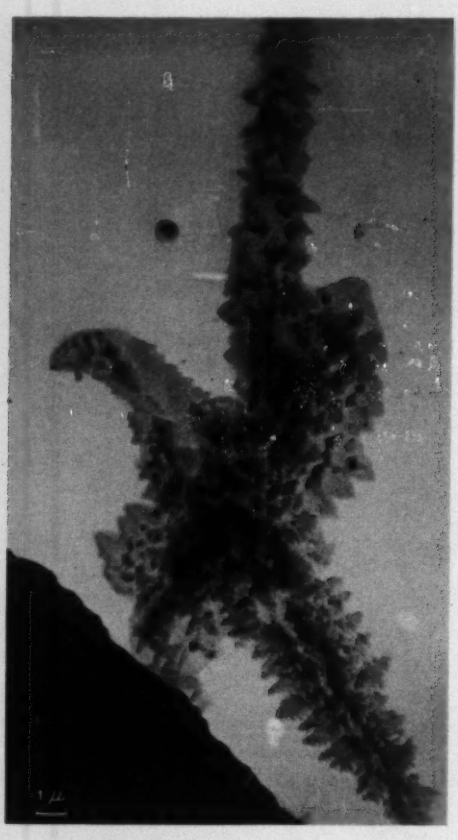


Fig. 3 — Bright-field micrograph of interpenetrant twin growth on (110) and (120) planes.

ten appear to be parallel to crystallographic planes. Certain patterns appear so consistently that it is difficult to ascribe them to mere accident. In the dark field of Fig. 6 some of the prominent loops in the upper part of the figure have sides which are parallel to (110) and (310) planes.

The dark field of Fig. 7 contains extinction contours that appear to be perfectly linear over distances of 5 to 10μ. Some of the crystallographic planes involved are illustrated in Fig. 8. There is one lamella which is bounded at a corner by (530) and ($\bar{5}$30) faces, while another lamella is bounded by (540) and ($\bar{5}$40) faces. These particular planes have been observed in a number of instances, either forming a corner or an edge

of a crystal aggregate. Thus, in the bright field of Fig. 9 there is a horizontal edge on the upper boundary which is parallel to ($\bar{5}40$).

Many diffraction patterns were recorded of crystal aggregates. Fine structure was invariably detected in the reflections if they were not exposed too heavily. The reflections appear at a casual glance to be more or less circular patches. Closer examination reveals a distinct mosaic structure in both radial and tangential directions. It is quite possible for moiré patterns to arise from tilts as well as rotations, and from combinations of both.[13] Tilts are likely to be operative in the cases of widely spaced fringes of 1000 angstroms or so.

III. CHAIN FOLDING — GENERAL CONSIDERATIONS

The types of chain folding discussed by Reneker and Geil[11] are given a simple representation in Fig. 10. A small rectangular patch of crystal is viewed along the *c* axis. The *a* axis is horizontal and the *b* axis vertical in the plane of the figure. (This crystallographic orientation is used throughout the remaining figures in this paper.) The molecular chains are normal to the plane of the figure. The molecules themselves are not indicated. The folds are situated in (110) planes. The distance that is

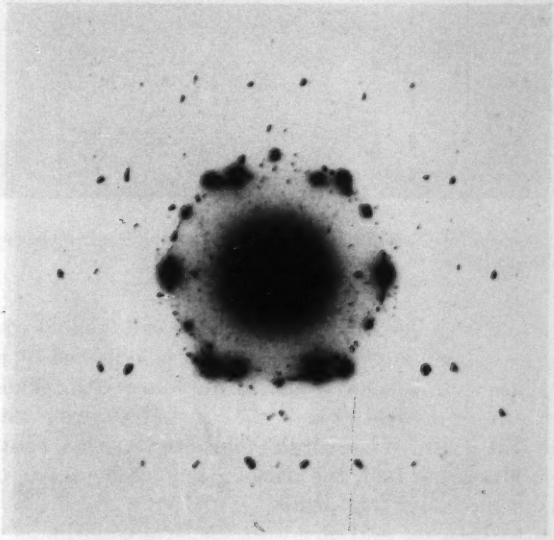


Fig. 4 — Electron diffraction pattern of crystals twinned on (110) and (120) planes.

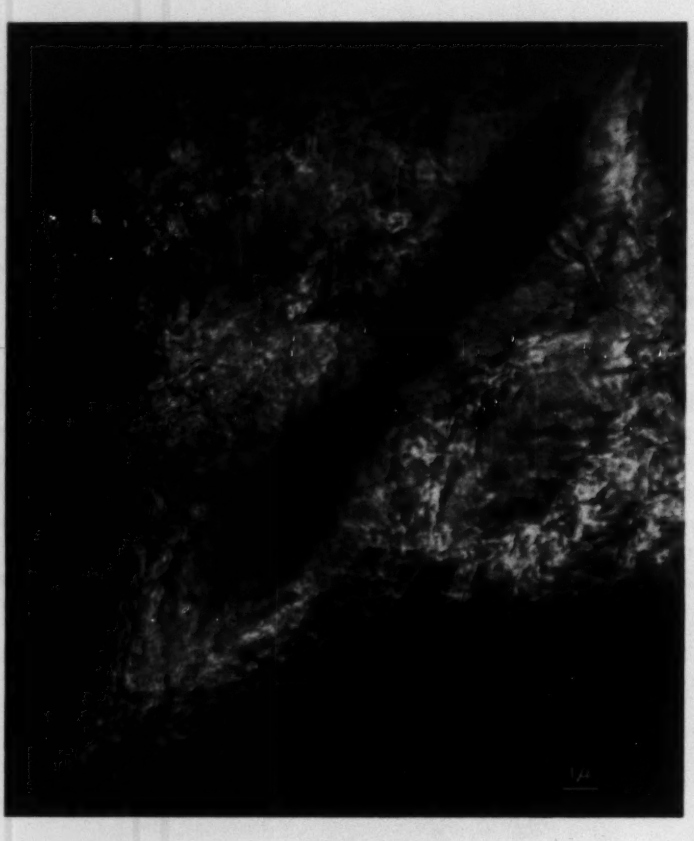


Fig. 5 — Electron interference effects in dark-field micrograph of crystal aggregate. Note the profusion of extinction lines and curves.

bridged by a fold from one chain to another we call the fold width. When a fold is on the upper surface of the crystal it is indicated by a solid line in a (110) plane with a length equal to the fold width. When a fold is on the lower surface there is a blank space in a (110) plane with a length equal to the fold width. When all the folds are identical, containing say three carbon atoms per fold, the packing requirements put restrictions on the folds in successive fold planes.

According to Reneker and Geil,[11] in the arrangement of Fig. 10(a) the folds in successive fold planes must be displaced by nc. For $n = 0$, the folds form upper and lower crystal surfaces which are tangent to

(001) and (00$\bar{1}$). For $n = 1$, if the displacement is always in the same direction, the folds form upper and lower crystal surfaces which are tangent to (111) and ($\bar{1}\bar{1}\bar{1}$), or tangent to (11$\bar{1}$) and ($\bar{1}\bar{1}$1). In the arrangement of Fig. 10(b) the folds in successive fold planes must be displaced by $(n + \frac{1}{2})c$. For $n = 0$, if the displacement is always in the same direction, the upper and lower crystal surfaces are tangent to (112) and ($\bar{1}\bar{1}\bar{2}$), or tangent to (11$\bar{2}$) and $\bar{1}\bar{1}$2). For $n = 1$, if the displacement is always in the same direction, the upper and lower crystal surfaces are tangent to (332) and $\bar{3}\bar{3}\bar{2}$), or tangent to (33$\bar{2}$) and ($\bar{3}\bar{3}$2). Reneker and Geil also suggest that, if all the folds are not identical—for example, if there is an

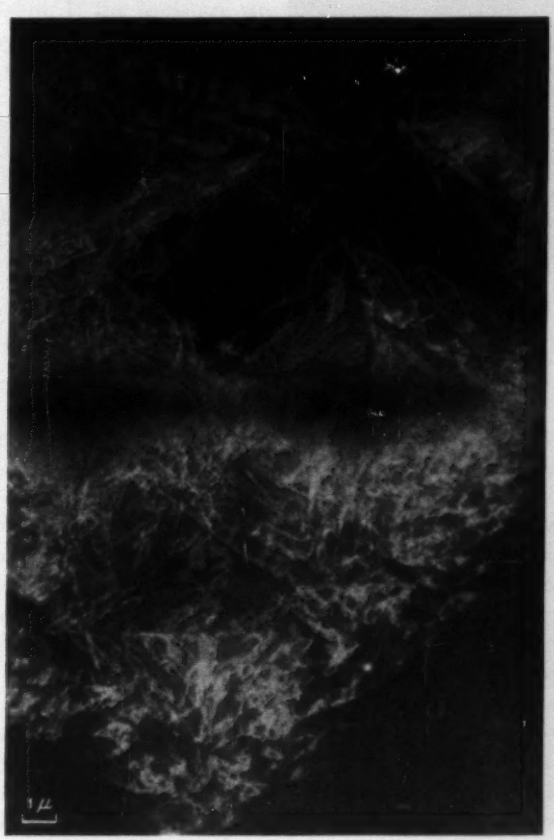


Fig. 6 — Dark-field micrograph. Some of the loops in the upper part of the figure have sides that are parallel to (110) and (310) planes.

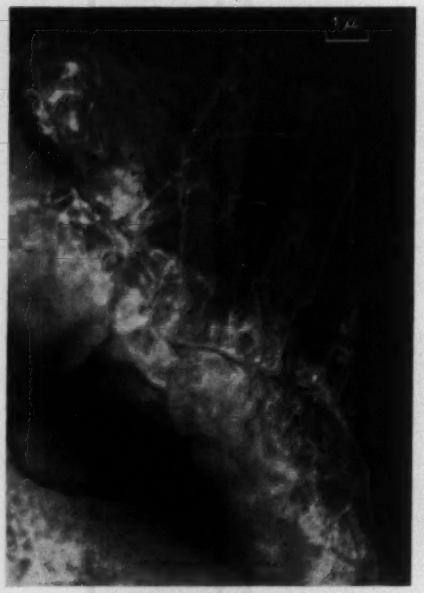


Fig. 7.

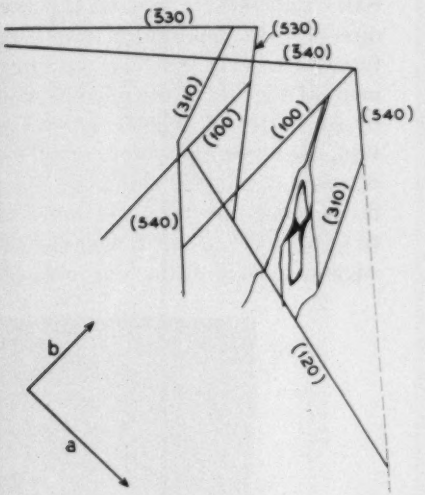


Fig. 8.

Fig. 7 — Dark-field micrograph with linear extinction lines along well-defined directions.
Fig. 8 — Identification of some of the crystallographic planes observed in Fig 7.

alternation between three atom folds and five atom folds in successive fold planes—then surfaces tangent to (001) are possible.

In the remainder of this paper attention will be directed to two-dimensional aspects of chain folding. However, it should be clear from the preceding paragraph that whenever there is a change from one folding structure to another the crystal surfaces formed will generally change their inclination to the (001) plane. It is essential to develop some kind of notation to describe the various chain folding arrangements. The notation (110)[010] describes the folding in Fig. 10(a). The folds lie in (110) planes. The shortest vector displacement from a fold in one plane to an equivalent fold in the next plane is **b**, or, in our notation, [010]. There is another orientation of the Fig. 10(a) structure which can be written $(1\bar{1}0)[010]$. The symbol $(110)[\frac{1}{2}\frac{1}{2}0]$ describes the folding in Fig. 10(b). Here the shortest vector displacement from one fold to an equivalent in the next plane is $\mathbf{a}/2 + \mathbf{b}/2$, or $[\frac{1}{2}\frac{1}{2}0]$. There is obviously another orientation of the Fig. 10(b) structure, which can be written $(1\bar{1}0)[\frac{1}{2}\frac{1}{2}0]$.

In principle, a crystal formed of (110)[010] folds might have a different melting point from a crystal formed of (110)[$\frac{1}{2}\frac{1}{2}$0] folds. In subsequent discussion we shall refer to such different arrangements as different phases. (The thermodynamicist, of course, would insist on additional criteria to define different phases.) The arrangements (110)[$\frac{1}{2}\frac{1}{2}$0] and (1$\bar{1}$0)[$\frac{1}{2}\bar{\frac{1}{2}}$0] do not represent different phases since they are mirror images of each other.

The observations of Bassett, Frank and Keller[8] on six-sector crystals with (100) faces indicate another type of fold structure. The most reasonable interpretation is that there are folds lying in planes parallel to the (100) faces. If the folds are situated in (200) planes the fold width increases by about 10 per cent relative to folds in (110) planes. Fig. 11 illustrates the folding structures that might be expected for this type

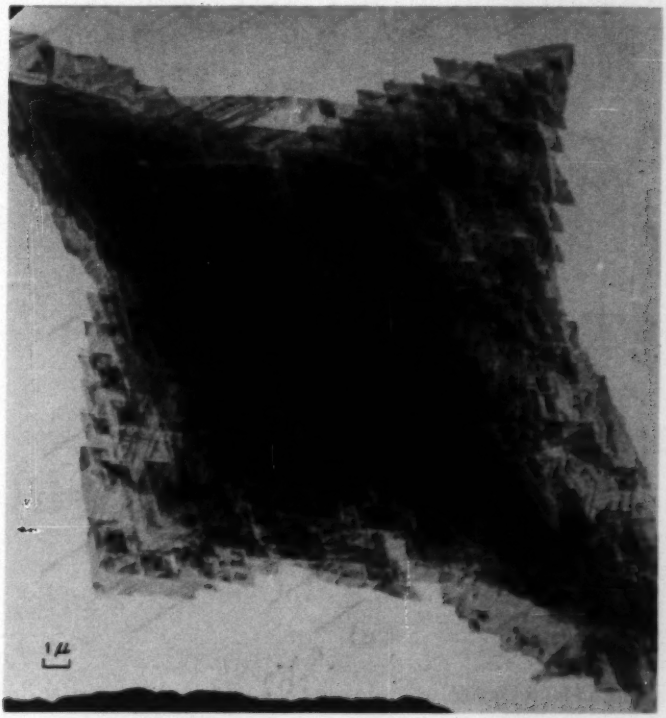


Fig. 9 — Bright field micrograph. Note horizontal edge on upper boundary which is parallel to (540).

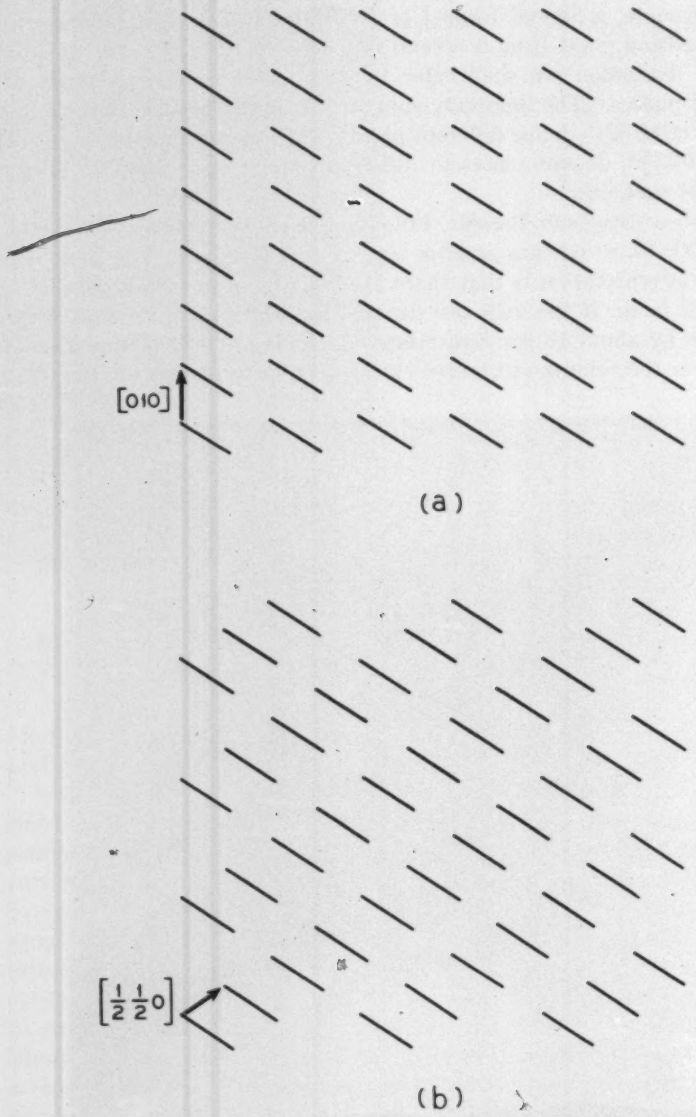


Fig. 10 — (a) (110)[010] fold structure; (b) (110)[$\frac{1}{2}\frac{1}{2}$0] fold structure. Crystal axes not illustrated but the a axis is horizontal, the b axis is vertical. This orientation is used in all the figures that follow.

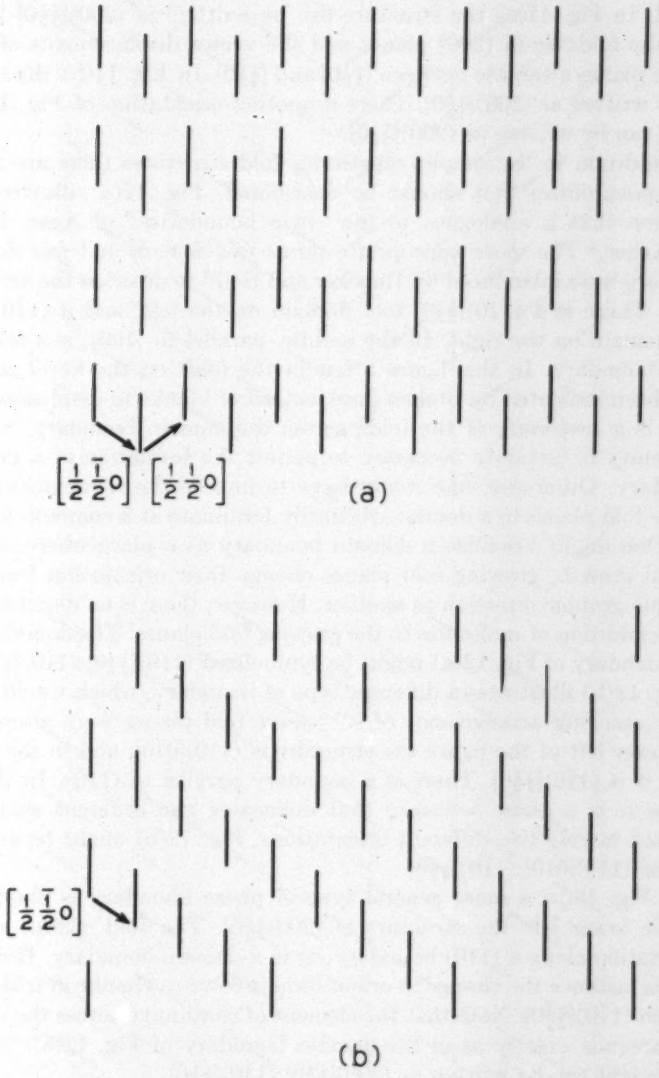


Fig. 11 — (a) (200)$[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]$ fold structure; (b) (200)$[\frac{1}{2}\bar{\frac{1}{2}}0]$ fold structure.

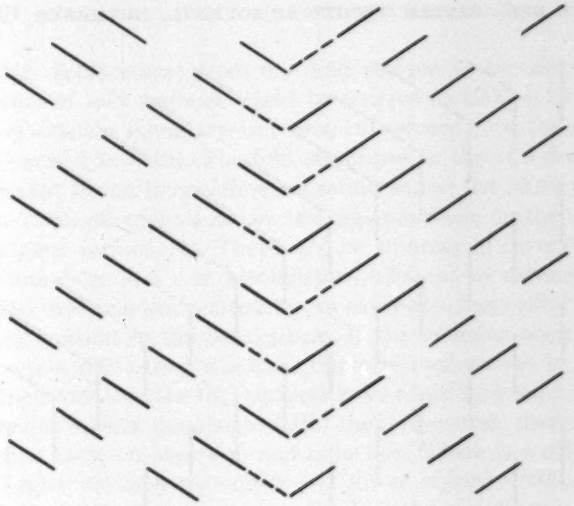
of fold. In Fig. 11(a) the structure can be written as $(200)[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]$, since the folds lie in (200) planes and the vector displacements of successive planes alternate between $[\frac{1}{2}\frac{1}{2}0]$ and $[\frac{1}{2}\bar{\frac{1}{2}}0]$. In Fig. 11(b) the structure is written as $(200)[\frac{1}{2}\bar{\frac{1}{2}}0]$. There is another orientation of Fig. 11(b), which can be written as $(200)[\frac{1}{2}\frac{1}{2}0]$.

In addition to the simple, repetitious fold structures there are many other possibilities that should be considered. Fig. 12(a) illustrates a situation that is analogous to the "twin boundaries" of Agar, Frank and Keller.[6] The more appropriate terms *fold domain* and *fold domain boundary* were introduced by Reneker and Geil[11] to describe the arrangement. There is a $(110)[\frac{1}{2}\frac{1}{2}0]$ fold domain on the left, and a $(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]$ fold domain on the right. In the middle, parallel to (100), is a fold domain boundary. In this figure a few of the folds on the lower surface have been indicated by broken lines instead of blanks to emphasize that there is a *continuity* of the folds across the domain boundary. Such a continuity is certainly necessary to permit the formation of a domain boundary. Otherwise, one would have to impose the requirement that all the fold planes in a doman arbitrarily terminate at a common boundary. One might visualize a domain boundary as a place where, during crystal growth, growing fold planes change their orientation from one crystallographic direction to another. However, there is no discontinuity in the addition of molecules to the growing fold planes. The domains and the boundary of Fig. 12(a) might be symbolized $(110)[\frac{1}{2}\frac{1}{2}0]:(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]$.

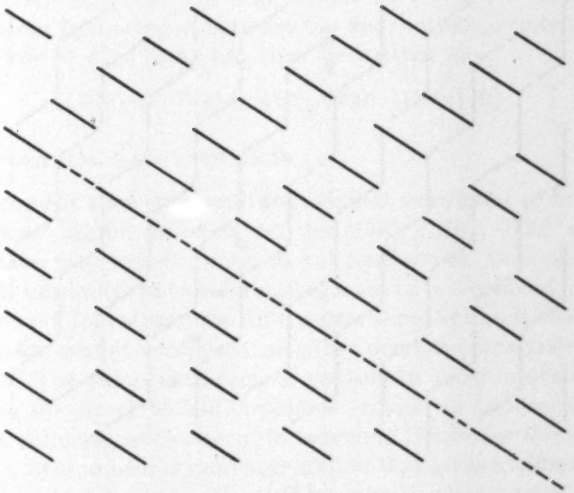
Fig. 12(b) illustrates a different type of boundary, which would occur if the stacking arrangement of successive fold planes were altered. In the lower left of the figure the structure is (110)[010], and in the upper right it is $(110)[\frac{1}{2}\frac{1}{2}0]$. There is a boundary parallel to (110). In this instance it is a *phase boundary* that delineates two different structures and not merely two different orientations. Fig. 12(b) might be symbolized as $(110)[010]:(110)[\frac{1}{2}\frac{1}{2}0]$.

In Fig. 13(a) a more general type of phase boundary is illustrated. In the lower left the structure is $(200)\frac{1}{2}\bar{\frac{1}{2}}0]$. The fold planes change orientation along a (110) boundary, as in a domain boundary. However, in this instance the change in orientation involves a change in fold structure to $(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]$. Note that the element of continuity across the boundary prevails exactly as in the domain boundary of Fig. 12(a). The arrangement can be written as $(200)[\frac{1}{2}\bar{\frac{1}{2}}0]:(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]$.

In Fig. 13(b) a more complex possibility is suggested. The folds are not confined to a fold plane but lie in a corrugated surface, the *fold surface*. Successive folds in the fold surface alternate between (200) and (110) planes. Hence the fold surface might be symbolized as $(200)\cdot(110)$.



(a)



(b)

Fig. 12 — (a) (110)[$\frac{1}{2}\frac{1}{2}$0]:(1$\bar{1}$0)[$\frac{1}{2}\frac{\bar{1}}{2}$0] domain boundary; (b) (110)[010]:(110)[$\frac{1}{2}\frac{1}{2}$0] phase boundary.

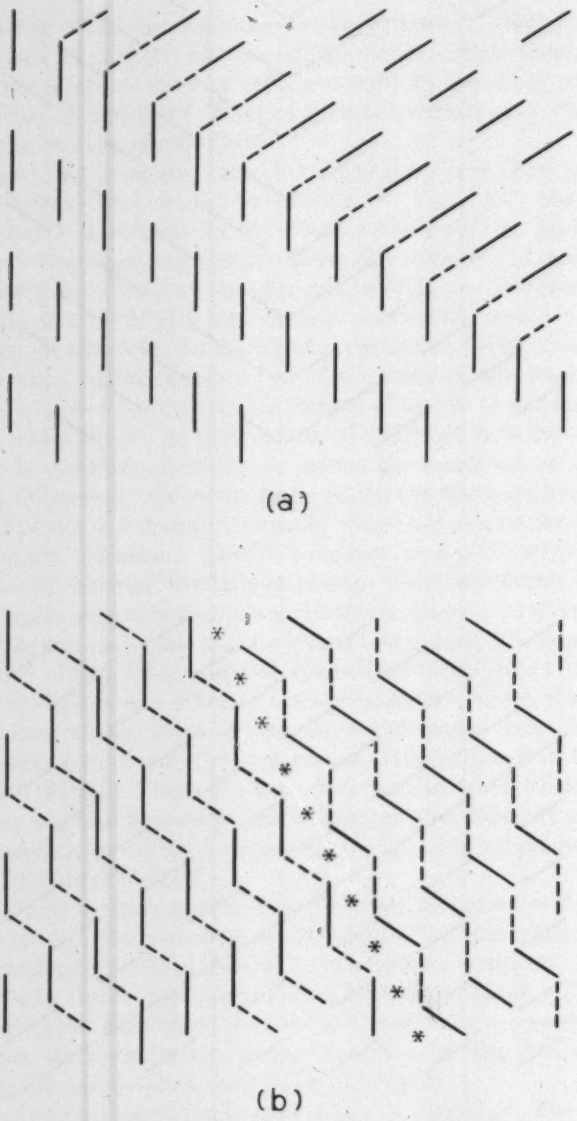


Fig. 13 — (a) $(200)[\frac{1}{2}\bar{1}0]:(1\bar{1}0)[\frac{1}{2}\bar{1}0]$ phase boundary; (b) $(200)\cdot(110)[\frac{1}{2}\bar{1}0]:[\frac{1}{2}\bar{1}0]i:(200)\cdot(110)[\frac{1}{2}\bar{1}0]$ domain boundary.

The identity displacement from one fold surface to the next is $[\frac{1}{2}\frac{1}{2}0]$. The structure of fold surfaces might be written as $(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]$. A new type of domain boundary has been introduced into the middle of the figure parallel to (310). The fold structures in the two domains are identical except for an inversion *or* reflection across the (001) plane. On the left the (200) plane folds are on the upper surface, on the right they are on the lower surface, etc. There *may* be an array of inversion points along the boundary, and this possibility is indicated by asterisks in Fig. 13(b). If the domains are related by an inversion they will both have the same inclination to the basal plane. If the domains are related by reflection across (001) they will have opposite inclinations to the basal plane. In the event that the fold surfaces have a stacking sequence which forms a crystal surface tangent to (001) then, of course, there would be no distinction between inversion and reflection. There is a difference in surface energies between the upper and lower crystal surfaces. Whatever the difference in energies may be in the left-hand domain it will be reversed in the right-hand domain. Also, the binding energy between the two fold surfaces that comprise the domain boundary should be different from the binding energy between adjacent fold surfaces within a domain. The relation between the two fold surfaces at the boundary can be written as $[\frac{1}{2}\frac{1}{2}0]i$. The fold surface on the left is translated by $[\frac{1}{2}\frac{1}{2}0]$, followed by inversion between top and bottom, symbolized by i. The structure of Fig. 13(b) can then be written as

$$(200)\cdot(110)[\tfrac{1}{2}\tfrac{1}{2}0]:[\tfrac{1}{2}\tfrac{1}{2}0]i:(200)\cdot(110)[\tfrac{1}{2}\tfrac{1}{2}0].$$

IV. CHAIN FOLDING AND TWINNING

At the present time it appears that crystal aggregates of linear polyethylene may exhibit twinning on the (110), (310), (530) and (120) planes. There may well be other, as yet unobserved, twin planes. This remarkable versatility in twin formation must have a rational interpretation in terms of fold structures. In the preceding section it was required that the folds exhibit continuity at either domain boundaries or phase boundaries. The same condition of continuity must prevail at twin boundaries. However, the fold structures provide an additional variable that is not ordinarily encountered in twinning. Hence, for this discussion we define a twin to be a crystal aggregation that gives a diffraction pattern which is based on two identical reciprocal lattices with a definite crystallographic relation between them. This definition is used because we will consider aggregations in which there are two space lattices related by a twin law. The crystal structure with respect to the chains

and chain packing is identical on the two space lattices, but the fold structure is different on the two space lattices. The boundary between regions of such an aggregation will always be a twin boundary with respect to the space lattice and crystal structure inside the folds. However, with respect to the surface folds the boundary may be a domain boundary (a true twin boundary) or a phase boundary.

In the preceding section it was shown that boundaries in the fold structure may be formed either by an alteration in the stacking sequence of fold planes or by a change in orientation of the fold planes. Investigation of (110) twinning shows that each of these mechanisms can produce both domain boundaries and phase boundaries in the fold structure. Thus, there are four types of twinning to consider across a (110) boundary.

In Fig. 14 is illustrated an example of a change in stacking sequence to form a domain boundary. Expanding our previous notation, this can

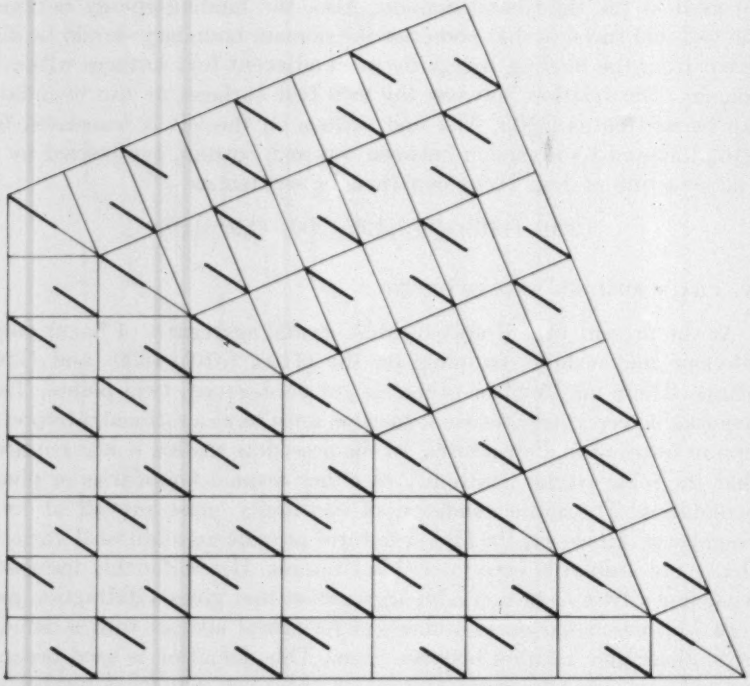


Fig. 14 — (110)[$\bar{1}\bar{1}$0]:(110):(110)[$\bar{1}\bar{1}$0] domain boundary in (110) twin.

be written as (110)[$\frac{1}{2}\frac{1}{2}$0]:(110):(110)[$\frac{1}{2}\frac{1}{2}$0], the middle (110) entry being added to indicate that the space lattices are twinned across the (110) plane.

In Fig. 15 the stacking sequence is altered to form a phase boundary. In this case the appropriate symbolism is (110)[010]:(110):(110)[$\frac{1}{2}\frac{1}{2}$0].

In Fig. 16 the fold plane changes orientation to form a domain boundary. The fold structure can be written as (1$\bar{1}$0)[$\frac{1}{2}\frac{\bar{1}}{2}$0]:(110):(1$\bar{1}$0)[$\frac{1}{2}\frac{\bar{1}}{2}$0].

In Fig. 17 the fold plane changes orientation to form a phase boundary. In this case the notation is somewhat more complex because of the stacking sequences: (200)[$\frac{1}{2}\frac{1}{2}$0]·[$\frac{1}{2}\frac{\bar{1}}{2}$0]:(110):(1$\bar{1}$0)[0$\bar{1}$0]·[$\frac{1}{2}\frac{\bar{1}}{2}$0].

All told we have found 11 distinct fold structures across a (110) twin plane when we confine the original or left-hand member of the twin to the simple stacking sequences described in Section III. To save space, these structures are summarized in Table I in fold structure notation. It is possible to reconstruct a line drawing of any of these structures from the notation and the condition of continuity of the folds.

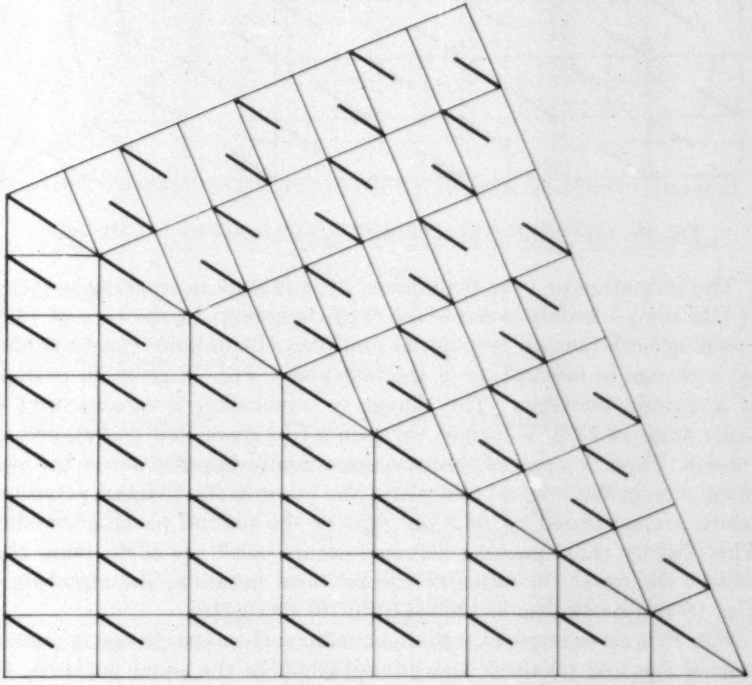


Fig. 15 — (110)[010]:(110):(110)[$\frac{1}{2}\frac{1}{2}$0] phase boundary in (110) twin.

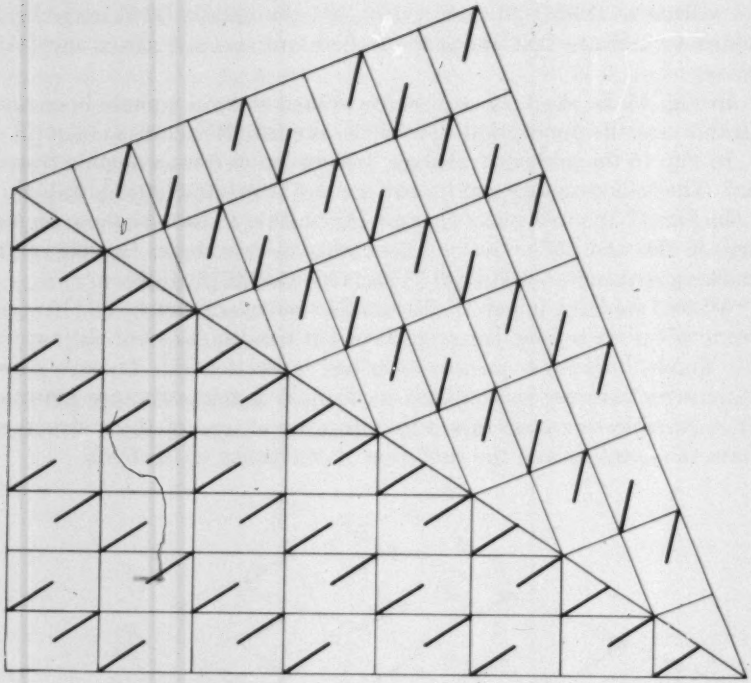


Fig. 16 — $(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]:(110):(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]$ domain boundary in (110) twin.

The formation of twin boundaries by a change in stacking sequence of fold planes is only possible for (110) twinning. In the case of (310) twinning both domain boundaries and phase boundaries can be formed by a change in orientation of the fold plane. Fig. 18 gives an example of a domain boundary. The change in orientation is extreme and an acute angle of 53°8′ is formed between a fold plane and its twinned extension. There is a pair of chains symmetrically disposed across the twin plane, one in the original fold plane, the other in the twinned extension, which are separated by 98.8 per cent of the normal packing distance. This slightly short packing distance occurs in all the (310) twins that will be discussed. In terms of the previous notation, the structure of Fig. 18 can be written as $(200)[\frac{1}{2}\bar{\frac{1}{2}}0]:(310):(200)[\frac{1}{2}\bar{\frac{1}{2}}0]$.

Fig. 19 is an example of a phase boundary. Here the change in orientation of the fold planes is almost negligible. In the usual notation, the structure is written as $(110)[010]:(310):(200)[\frac{1}{2}\bar{\frac{1}{2}}0]$.

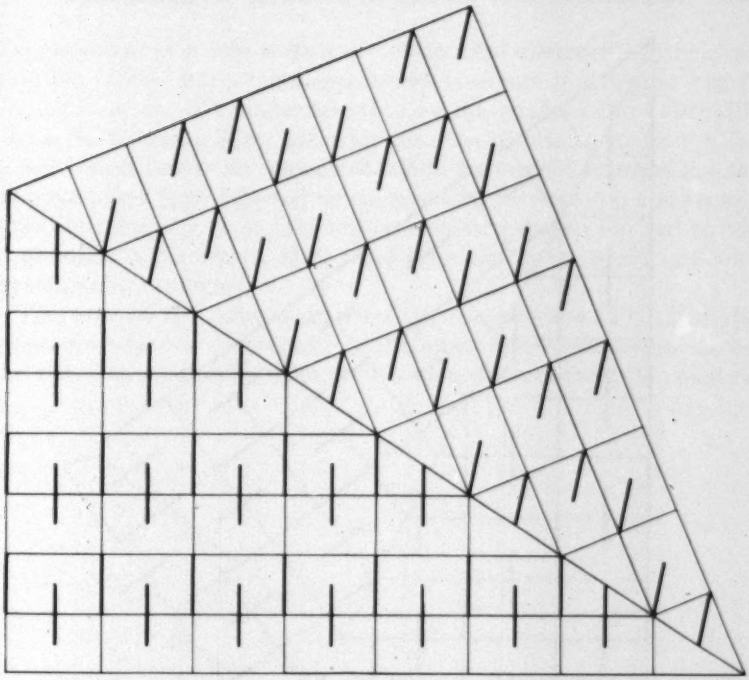


Fig. 17 — $(200)[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]:(110):(1\bar{1}0)[0\bar{1}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]$ phase boundary in (110) twin.

TABLE I—SOME GEOMETRICALLY POSSIBLE FOLD STRUCTURES IN (110) TWINNING

Change in Stacking Sequence of Fold Planes	
To form domain boundary	1. $(110)[\frac{1}{2}\frac{1}{2}0]:(110):(110)[\frac{1}{2}\frac{1}{2}0]$ 2. $(110)[010]:(110):(110)[010]$
To form phase boundary	3. $(110)[010]:(110):(110)[\frac{1}{2}\frac{1}{2}0]$
Change in Orientation of Fold Planes	
To form domain boundary	4. $(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]:(110):(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]$ 5. $(1\bar{1}0)[010]:(110):(1\bar{1}0)[010]$ 6. $(200)[\frac{1}{2}\frac{1}{2}0]:(110):(200)[\frac{1}{2}\frac{1}{2}0]$ 7. $(200)[\frac{1}{2}\frac{1}{2}0]:(110):(200)[\frac{1}{2}\bar{\frac{1}{2}}0]$ 8. $(200)[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]:(110):(200)[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]$
To form phase boundary	9. $(200)[\frac{1}{2}\frac{1}{2}0]:(110):(1\bar{1}0)[010]$ 10. $(200)[\frac{1}{2}\frac{1}{2}0]:(110):(1\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]$ 11. $(200)[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]:(110):(1\bar{1}0)[0\bar{1}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]$

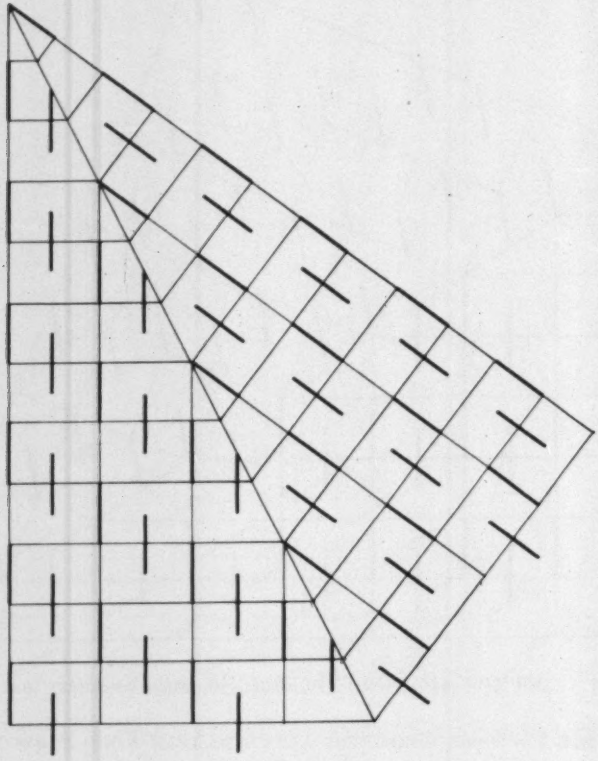


Fig. 18 — (200)[$\frac{1}{2}\bar{1}$0]:(310):(200)[$\frac{1}{2}\bar{1}$0] domain boundary in (310) twin.

Fig. 20 is still another phase boundary, which involves a new detail. Alternate fold planes contain a fold which does not terminate at the twin plane. Instead, this fold straddles the twin plane and has a fold width that is 98.8 per cent of the width of a (110) fold. The previous notation suffices to cover this detail as long as it is understood that the condition of continuity must be satisfied, and the structure is written as (1$\bar{1}$0)[010]:(310):(1$\bar{1}$0)[$\frac{1}{2}\bar{1}$0].

Nine distinct fold structures have been found when we confine the original or left-hand member of the twin to the simple stacking sequences. These structures are summarized in Table II.

An entirely different possibility for twin boundaries arises if we consider structures based on fold surfaces instead of fold planes Some of

the properties of a fold surface structure were discussed in connection with Fig. 13(b), and a more generalized structure is shown in Fig. 21. The twin on the left can be described by the symbol $(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]i$, where the notation $[\frac{1}{2}\frac{1}{2}0]i$ indicates that the displacement from a fold in one fold surface to an equivalent fold in the next fold surface involves the translation $[\frac{1}{2}\frac{1}{2}0]$ followed by an inversion between top and bottom. Since the folds of a given fold width alternate between top and bottom at alternate fold surfaces, there is no difference in the upper and lower crystal surface energies.

The twin on the right is described by the notation $(200)(110)[\frac{1}{2}\frac{1}{2}0]$, where now there *is* a difference in the upper and lower crystal surface energies. The relation between the two structures at the twin plane can

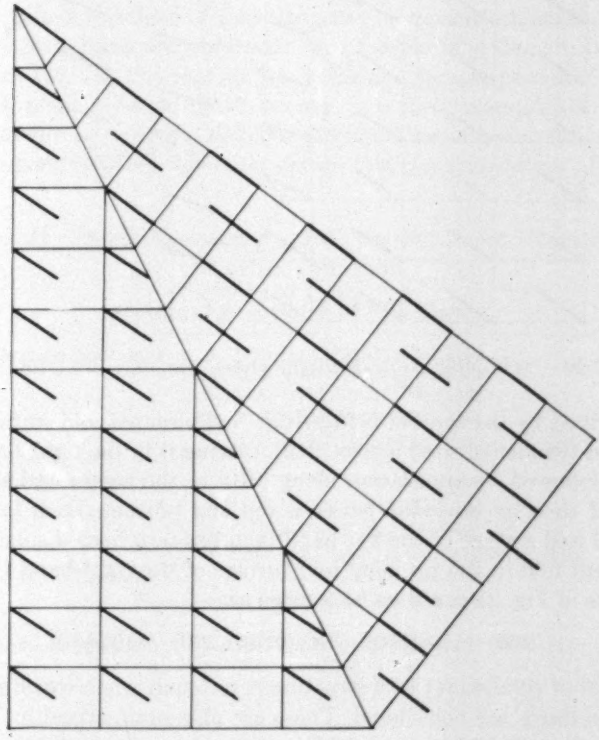


Fig. 19 — (110)[010]:(310):(200)$[\frac{1}{2}\bar{\frac{1}{2}}0]$ phase boundary in (310) twin.

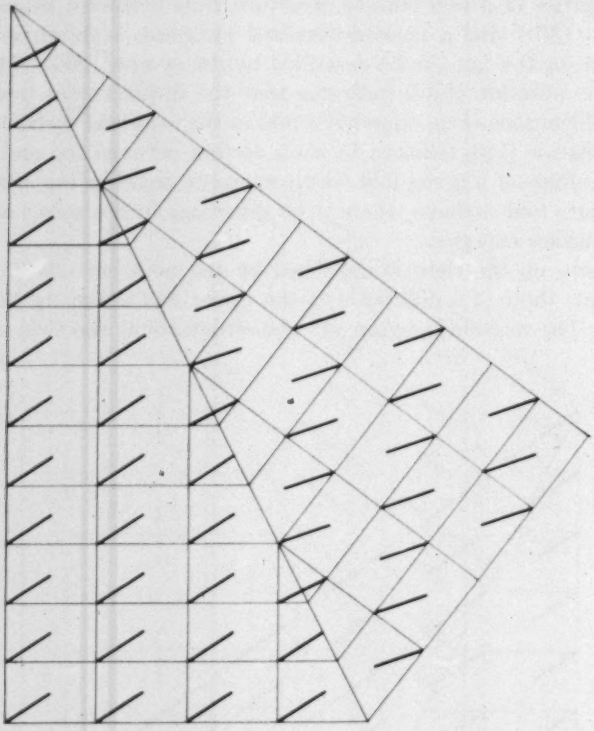


Fig. 20 — $(1\bar{1}0)[010]:(310):(1\bar{1}0)[\frac{1}{2}\frac{1}{2}\bar{0}]$ phase boundary in (310) twin.

be described by the symbol $(310)[0\bar{1}0]i$. The terminal fold surface in the left-hand twin is reflected across (310), placing it in the right-hand twin. This is followed by translation along $[0\bar{1}0]$ in the right-hand space lattice, and then by inversion between top and bottom. Each fold in the terminal fold surface of the left-hand twin has now been displaced to an equivalent fold in the terminal fold surface of the right-hand twin. The structure of Fig. 21 can then be written as

$$(200)\cdot(110)[\tfrac{1}{2}\tfrac{1}{2}0]i:(310)[0\bar{1}0]i:(200)\cdot(110)[\tfrac{1}{2}\tfrac{1}{2}0].$$

A total of six distinct fold structures are found when transformations of fold surfaces are considered. These are also summarized in Table II, making a total of 15 distinct fold structures across a (310) twin plane.

For (530) twinning, only one example has been found of a boundary

formed by a change in orientation of a fold plane. This example is illustrated in Fig. 22. It is similar to the structure of Fig. 20 in that it is a phase boundary and two thirds of the fold planes contain a fold which straddles the twin plane. The widths of these straddle folds are 145 per cent and 98.7 per cent of the width of a (110) fold. The latter slightly short packing distance occurs in all the (530) twins that will be discussed. In contrast to the structure of Fig. 20, there are pairs of vacant lattice sites which straddle the (530) boundary at regular intervals. These vacancies are inevitable if, in addition to the condition of continuity, we require that neighboring chains cannot be appreciably closer together than the width of a (110) fold. The structure of Fig. 22 can be written as $(200)[\frac{1}{2}\bar{\frac{1}{2}}0]:(530):(200)[\frac{1}{2}\bar{\frac{1}{2}}0]\cdot 2[\frac{1}{2}\frac{1}{2}0]$, where the notation $[\frac{1}{2}\bar{\frac{1}{2}}0]\cdot 2[\frac{1}{2}\frac{1}{2}0]$ indicates that a displacement of $[\frac{1}{2}\bar{\frac{1}{2}}0]$ is followed by two successive displacements of $[\frac{1}{2}\frac{1}{2}0]$.

When transformations of fold surfaces are considered, six more (530) boundary structures are obtained. An example of a domain boundary is given in Fig. 23. The fold surface is defined by a sequence of one fold in a (200) plane followed by three folds in a (110) plane. Thus, the fold surface might be written as $(200)\cdot 3(110)$. The transformation across the boundary involves reflection across (530), a translation of $[\bar{1}10]$ in

TABLE II — SOME GEOMETRICALLY POSSIBLE FOLD STRUCTURES IN (310) TWINNING

	Change in Orientation of Fold Planes
To form domain boundary	1. $(110)[\frac{1}{2}\frac{1}{2}0]:(310):(110)[\frac{1}{2}\frac{1}{2}0]$
	2. $(110)[010]:(310):(110)[010]$
	3. $(200)[\frac{1}{2}\frac{1}{2}0]:(310):(200)[\frac{1}{2}\frac{1}{2}0]$
	4. $(200)[\frac{1}{2}\frac{1}{2}0]:(310):(200)[\frac{1}{2}\frac{1}{2}0]$
	5. $(200)[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]:(310):(200)[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]$
To form phase boundary	6. $(110)[\frac{1}{2}\frac{1}{2}0]:(310):(200)[\frac{1}{2}\frac{1}{2}0]$
	7. $(110)[010]:(310):(200)[\frac{1}{2}\bar{\frac{1}{2}}0]$
	8. $(1\bar{1}0)[010]:(310):(1\bar{1}0)[\frac{1}{2}\frac{1}{2}0]$
	9. $(200)[\frac{1}{2}\frac{1}{2}0]\cdot[\frac{1}{2}\bar{\frac{1}{2}}0]:(310):(110)[010]\cdot[\frac{1}{2}\frac{1}{2}0]$
	Transformation of Fold Surfaces
To form domain boundary	10. $(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]:(310)[0\bar{1}0]:(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]$
	11. $(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]i:(310)[0\bar{1}0]:(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]i$
	12. $(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]:(310)[0\bar{1}0]i:(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]$
	13. $(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]i:(310)[0\bar{1}0]i:(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]i$
To form phase boundary	14. $(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]i:(310)[0\bar{1}0]:(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]$
	15. $(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]i:(310)[0\bar{1}0]i:(200)\cdot(110)[\frac{1}{2}\frac{1}{2}0]$

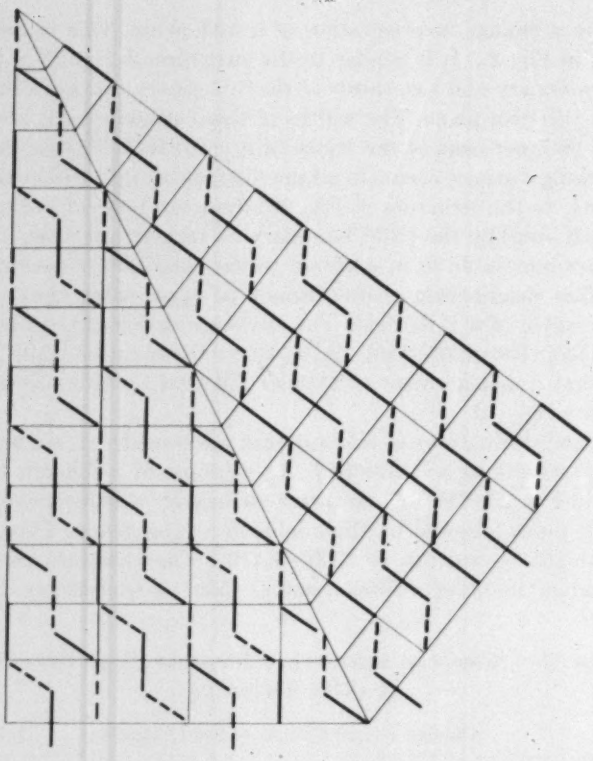


Fig. 21 — (200)·(110)[$\frac{1}{2}\frac{1}{2}$0]i:(310)[0$\bar{1}$0]i:(200)·(110)[$\frac{1}{2}\frac{1}{2}$0] phase boundary in (310) twin.

the right-hand space lattice, and inversion. Hence the structure of Fig. 23 can be written as (200)·3(110)[$\frac{1}{2}\frac{1}{2}$0]:(530)[$\bar{1}$10]i:(200)·3(110)[$\frac{1}{2}\frac{1}{2}$0]. There are four domain boundaries and two phase boundaries based on fold surfaces, exactly as for the (310) boundaries listed in the lower part of Table II.

There is good evidence in Fig. 7 for the existence of (530) crystal faces. The existence of fold surfaces seems quite likely, both in terms of observed morphology and observed twin relations. The experimental evidence for (120) twinning is perhaps somewhat weaker. However, if (120) twin boundaries do exist the possibilities are quite analogous to the (530) case. One phase boundary has been found based on a change in orientation of a fold plane. There are boundary folds whose widths are 21.5 per cent and 5 per cent larger than the width of a (110) fold. The struc-

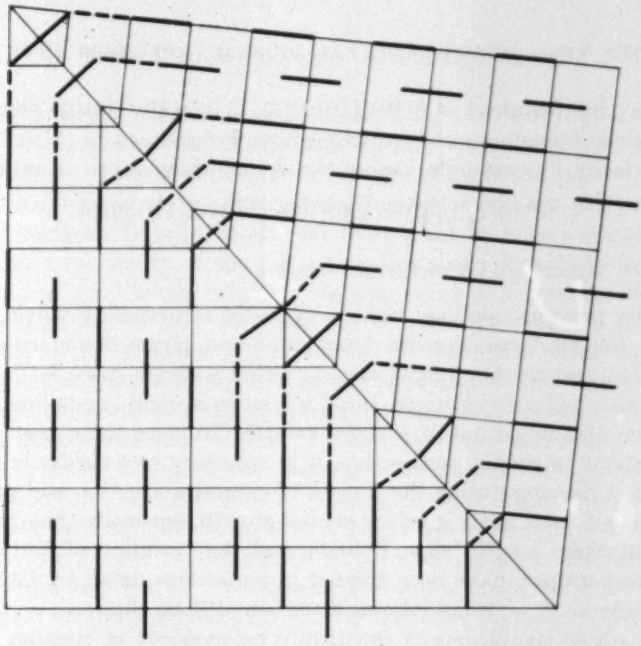


Fig. 22 — $(200)[\frac{1}{22}\bar{1}0]:(530):(200)[\frac{1}{22}\bar{1}0]\cdot 2[\frac{1}{22}10]$ phase boundary in (530) twin.

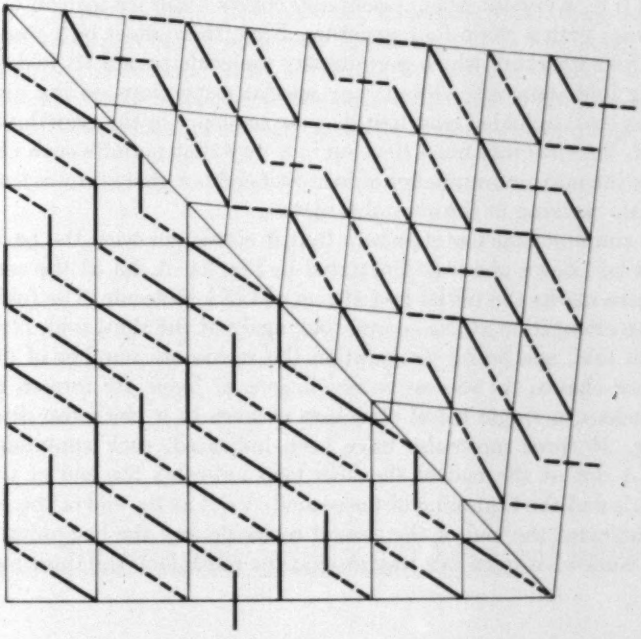


Fig. 23 — $(200)\cdot 3(110)[\frac{1}{22}10]:(530)[\bar{1}10]_i:(200)\cdot 3(110)[\frac{1}{22}10]$ domain boundary in (530) twin.

ture may be described as $(1\bar{1}0)[\frac{1}{2}\frac{1}{2}\bar{1}0]:(120):(1\bar{1}0)[\frac{1}{2}\frac{1}{2}\bar{1}0]\cdot 2[0\bar{1}0]$. There are four domain boundaries and two phase boundaries based on $(1\bar{1}0)\cdot 3(110)$ fold surfaces. For example, one of the domain boundaries is written as

$$(1\bar{1}0)\cdot 3(110)[010]:(120)[\bar{1}00]:(1\bar{1}0)\cdot 3(110)[010].$$

V. CHAIN FOLDING AND CRYSTAL GROWTH

In the previous sections various types of structures involving fold planes, fold surfaces, domain boundaries and phase boundaries have been discussed as they might arise in either a single space lattice or a pair of twinned space lattices. In all instances a small crystalline region has been abstracted out of a larger crystal. To place these geometrical abstractions in proper perspective, it is necessary to consider how fold structures develop during the process of crystal nucleation and growth. The three-dimensional aspects of crystal growth, especially the relatively constant chain length between folds, and the variation of fold length with temperature, have been treated in meticulous detail by Lauritzen and Hoffman.[14] In what follows attention will be directed to various factors whose significance is practically independent of lamellar thickness.

There are two basic ingredients in the formation of a two-dimensional crystal (i.e., a crystal which essentially covers a surface instead of filling a volume) with a chain fold structure. First, there must be a continuity of the fold structure which permits any molecule to add its folds to the existing fold structure without any special restrictions on the molecule as to its total unfolded length or its time sequence in the growth process. Second, the structure must develop in a way that permits each chain to possess the maximum number of nearest-neighbor chains consistent with the chain packing in the crystal structure.

The simplest lamellar structure that is consistent with the basic conditions laid down above is illustrated in Fig. 24. A dot at the center of the figure marks the initial fold at one end of a molecule. The fold plane changes orientation at the second fold, again at the third fold, fifth fold, seventh fold, and so on to maintain the maximum number of nearest-neighbor chains. As successive revolutions or loops are formed around the nucleus, a single spiral structure evolves in a clockwise direction. In Fig. 24 three molecules have been indicated, each containing 100 folds. A dot at the end of the fifth loop indicates the end of the first molecule and the beginning of the second. A dot at the end of the seventh loop indicates the end of the second molecule and the beginning of the third. Somewhat over half way around the ninth loop the third molecule

ends. The crystal contains four domains of the (110)[010] type, two domain boundaries of the (110)[010]:(1$\bar{1}$0)[010] type, and two domain boundaries of the (110)[010]:($\bar{1}$10)[010] type. This is the type of crystal that was described by Reneker and Geil[11] as forming either (001) or (111) crystal surfaces. It is apparent that there must be both a spiral structure and a continuity of the fold planes from one domain to the next. There is no good reason why the starting point of a crystal might not be a fold at some other point than the end of a molecule. In this case, a double spiral structure will be formed at the center. This will have consequences for the subsequent structure of the crystal which will be discussed below. Both single-spiral and double-spiral crystal nuclei were among the possibilities recognized by Lauritzen and Hoffman.

The structure of Fig. 24 provides a convenient model for some simple calculations that are not without physical insight. From the geometry of the fold structure, if the number of loops around the spiral at a given point in the growth be designated n, then the number of folds in the last or outer loop is $4(2n - 1)$. Assume a weight average molecular weight of 130,000 for the polymer; then an average molecule contains 9285 CH_2 units. Suppose there are 93 CH_2 units per chain plus fold, with 3 CH_2 units per fold and 90 CH_2 units per chain. Then with 2.53 angstroms along the c axis per pair of CH_2 units, the fold length is 114 angstroms. There are approximately 100 folds per molecule. With these assumptions the total number of molecules in the crystal at any stage of the crystal growth is determined by a count of the number of folds. In addition, the number of molecules in the outer five loops is $(2n/5) - 1$.

The rate of addition of molecules to the crystal will be proportional

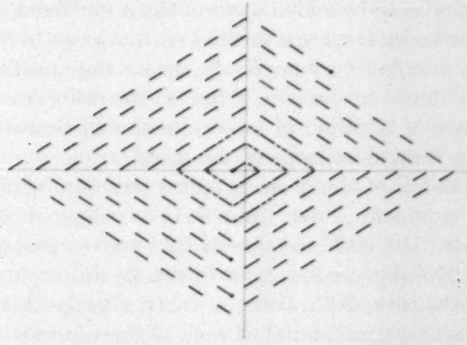


Fig. 24 — Single-spiral crystal containing (110)[010]-type fold domains. Dots indicate terminal points of three molecules included in figure.

to the average number of molecules in contact with the growing surface. This average number of molecules in turn is proportional to the area of the growing lateral faces of the crystal. The lateral area is proportional to the number of folds in the outer loop of the spiral. Thus, we can state that the average time for the formation of a fold on the growing surface is inversely proportional to $4(2n - 1)$, or

$$\text{time/fold} = \frac{K}{4(2n - 1)}.$$

Now assume that a crystal grows to 30μ in length along the a axis in 30 minutes. The number of loops encountered at various stages of the growth to 30μ size is listed in Table III, along with other time-independent quantities. The number of loops has grown to 20,480 when the crystal is about 30μ long. Now the constant K can be evaluated, since the time required for all the folds to form is

$$1800 \text{ seconds} = \sum_{n=1}^{n=20,480} \frac{K}{4(2n - 1)} [4(2n - 1)] = 20,480K,$$

or

$$K = \frac{1800}{20,480} = 0.08789 \text{ second.}$$

Based on the above assumptions, the average time per fold at various stages of growth is listed in Table III. Note that identical numerical values can be obtained by simply assuming that the same time interval is required to form each loop in the spiral. The quantity $5K = 0.439$ second is an interesting one, since it is the time required for an integral number of molecules to be added to form the outer five loops. The number of molecules varies from one in the first five loops to 8191 in the last five loops. The calculations completely ignore the question of the time required for the initial nucleation. Even so, the estimates of rate of fold formation or rate of addition of molecules should be reasonable for all but a few of the innermost loops of the spiral.

There is one detail of the growing spiral structure of Fig. 24 that requires further comment. Four times in every loop it is necessary to "round a corner." In each instance a fold has to project beyond the surface formed by the preceding loop. In Fig. 24 this occurs at the fourth, sixth, ninth, 12th, 16th, 20th, 25th, ... folds. Clearly there is an energy barrier that must be surmounted at each of these points. Lauritzen and Hoffman devoted considerable attention to this question.

However, there is more than one way in which a corner may

TABLE III — CRYSTAL GROWTH OF SINGLE-SPIRAL (110)[010] FOLD STRUCTURE

(Assuming Growth to 30μ Size in 30 Minutes)

Number of Loops, n	Number of Folds in Outer Loop, $4(2n-1)$	Number of Molecules in Crystal*	Length of Crystal	Number of Molecules in Outer 5 Loops,† $(2n/5)-1$	Average Time per Fold‡ (in seconds)
1	4	—	—	—	2.2×10^{-2}
2	12	—	—	—	7.3×10^{-3}
3	20	—	—	—	4.4×10^{-3}
4	28	—	—	—	3.1×10^{-3}
5	36	1	$6.7 \times 10^{-3}\mu$	1	2.4×10^{-3}
10	76	4	1.4×10^{-2}	3	1.2×10^{-3}
15	116	9	2.1×10^{-2}	5	7.6×10^{-4}
20	156	16	2.9×10^{-2}	7	5.6×10^{-4}
40	316	64	5.8×10^{-2}	15	2.8×10^{-4}
80	636	256	1.2×10^{-1}	31	1.4×10^{-4}
160	1,276	1,024	2.4×10^{-1}	63	6.9×10^{-5}
320	2,556	4,096	4.7×10^{-1}	127	3.4×10^{-5}
640	5,116	16,384	9.5×10^{-1}	255	1.7×10^{-5}
1,280	10,236	65,536	1.9	511	8.6×10^{-6}
2,560	20,476	262,144	3.8	1,023	4.3×10^{-6}
5,120	40,956	1,048,576	7.6	2,047	2.1×10^{-6}
10,240	81,916	4,194,304	1.5×10	4,095	1.1×10^{-6}
20,480	163,836	16,777,216	3.0×10	8,191	5.4×10^{-7}

* Number of folds in crystal = 100 (number of molecules in crystal) + 4, for all loops with $n > 5$.

† Outer five loops are formed in 0.439 second.

‡ Average time per molecule $\leqq$ 100 (average time per fold), for all loops with $n > 5$

be rounded. Consider the situation presented in Fig. 25. Again, there is a single spiral developing in a clockwise direction. At the fourth fold the corner has been rounded, not by adding another fold in a $(\bar{1}10)$ plane but by adding a fold in a (200) plane. It may require more energy to form a (200) fold. However, the altered arrangement of nearest-neighbor and next-nearest-neighbor chains will make other adjustments in the energy balance. There are dozens of configurations that conceivably might arise in forming the first two or three loops of a single spiral nucleus. In Fig. 25 the spiral has been constructed in such a way that each loop has one interruption by a (200) fold. The result of this is to produce a $[\frac{1}{2}\frac{1}{2}0]$ stacking sequence of (110) fold planes. One may protest that a "jog" is produced in every loop where the (200) fold occurs. In Fig. 26 an alternative way of producing a single spiral $(110)[\frac{1}{2}\frac{1}{2}0]$ fold structure is shown. There is no jog in the loop pattern, but a sequence of vacancies has been produced along the $(\bar{1}10)[\bar{\frac{1}{2}}\frac{1}{2}0]:(\bar{1}\bar{1}0)[\frac{1}{2}\bar{\frac{1}{2}}0]$ domain boundary. This may appear to be an even more unlikely structure than the preceding one. The interesting fact is that it is impossible to produce a

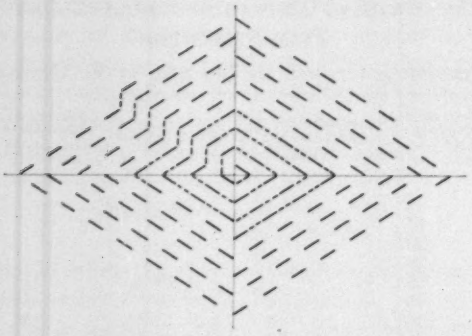


Fig. 25 — Single-spiral crystal containing (110)[$\frac{1}{2}\frac{1}{2}$0]-type domains. One (200) fold is required in every loop to maintain the stacking sequence.

single-spiral (110)[$\frac{1}{2}\frac{1}{2}$0] fold structure without introducing a (200) fold in each loop to maintain the [$\frac{1}{2}\frac{1}{2}$0] stacking sequence.

In Fig. 27 a clockwise double-spiral structure is illustrated. In this case a 115-fold molecule has been drawn at the center. The first fold in the nucleus is the 29th fold in the molecule. One branch of the double spiral contains 28 folds; the other contains 86 folds. As long as the two branches parallel each other, a (110)[$\frac{1}{2}\frac{1}{2}$0] structure is produced automatically. However, as soon as the long branch overlaps the short branch, the structure reverts to (110)[010]. If a second molecule adds to the crystal in single-spiral form the (110)[010] structure will continue, as illustrated in Fig. 27.

To produce an extensive double-spiral structure many coincidences

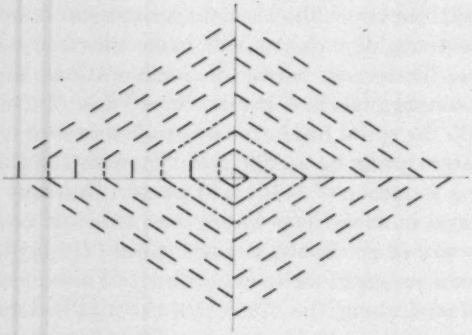


Fig. 26 — Single-spiral crystal containing (110)[$\frac{1}{2}\frac{1}{2}$0]-type domains. When the (200) folds are arrayed along the horizontal axis a vacancy is generated with each loop.

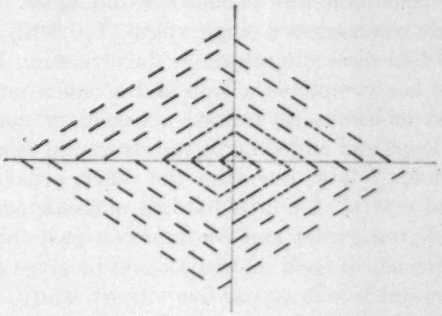


Fig. 27 — Double-spiral nucleus with first fold located 29 folds from one end of molecule. Monolayer nucleation and addition of a second molecule.

are necessary. The first fold in the nucleus must occur at the middle of a molecule. Later molecules must add as double monolayers with the first fold at the middle of the molecule. However, it is inconceivable that later molecules would not also add as two monolayers proceeding in both clockwise and counterclockwise directions. It seems very unlikely that an extensive (110)[$\frac{1}{2}\frac{1}{2}$0] structure can be produced from either single or double spirals, without introducing a systematic array of (200) folds into the structure as in Fig. 25.

Fig. 28 gives an indication of how complex and interwoven the relations between molecules may become, depending on their sequence in

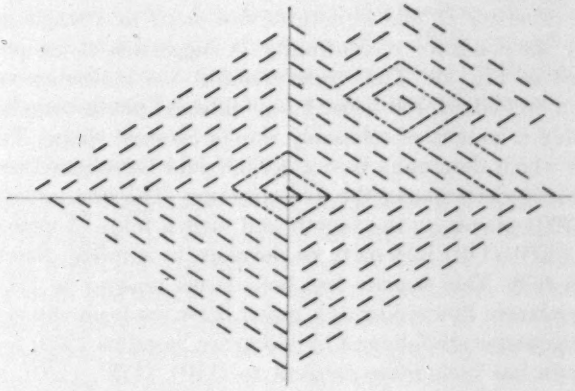


Fig. 28 — Single spiral (110)[010] structure, showing four molecules and spiral secondary nucleation of third molecule in time sequence.

time. There are four molecules included in the figure. At the center, a 100-fold molecule commences a single spiral (110)[010] structure. After five loops a 150-fold molecule continues the structure. By the time the second molecule has completed a loop and a half a new event occurs. A third 101-fold molecule has formed a secondary nucleus which has completed two loops and added on to the structural sequence. Now two monolayers of folds follow, one after the other, around the perimeter of the compound crystal. A fourth 100-fold molecule adds on where the 150-fold molecule terminates and continues on past the termination of the 101-fold molecule. If each molecule could be given a color, the perimeter of the crystal would appear like striped candy. Although an integral (110)[010] phase has been maintained throughout, the addition of the secondary nucleus is revealed by the presence of four new fold domains and six new domain boundaries.

If conditions are favorable for the presence of spiral secondary nuclei it is likely that they would be attracted to the innermost corners of the re-entrant faces. Hence, if a secondary nucleus adds to a sizable primary crystal, a whole string of secondary nuclei might be added, one after another. Such an array would lie in a virtually straight line if the time sequence were rapid. It is apparent that, as a crystal grows in size, whether from addition of secondary spirals or monolayers, traffic jams are bound to develop. Additions will occur in both clockwise and counterclockwise directions; there will be collisions with adjustments to make. Under such conditions one would expect a variety of fold surfaces and other space lattice orientations to become more probable as a crystal aggregate develops.

A great diversity of fold structures can occur in a single spiral and still satisfy the condition of continuity. A suggestion of the possibilities is indicated in Fig. 29. The figure contains ten molecules varying in length from 30 folds to 100 folds. Many different phase boundaries have arisen within a lamella of relatively simple external shape. The precise conditions which determine how a corner shall be rounded are the decisive factor in determining the fold structure. The first molecule forms folds in (200) planes at the fourth and eighth folds. A phase domain based on (200)·(110) fold surfaces develops in a radial direction from the fourth fold. This domain has boundaries tangent to (110), $(1\bar{3}0)$, and $(\bar{1}\bar{1}0)$ planes. Proceeding in a radial direction from the eighth fold, a pair of vacancies are followed by a domain based on (200) fold planes. This domain has boundaries tangent to (110), $(1\bar{1}0)$, $(\bar{1}\bar{1}0)$, and $(\bar{1}10)$ planes. Most of the area of the crystal is filled in with (110) fold planes. However, the stacking sequences are irregular because of the varying numbers of (200) folds that occur in the different loops.

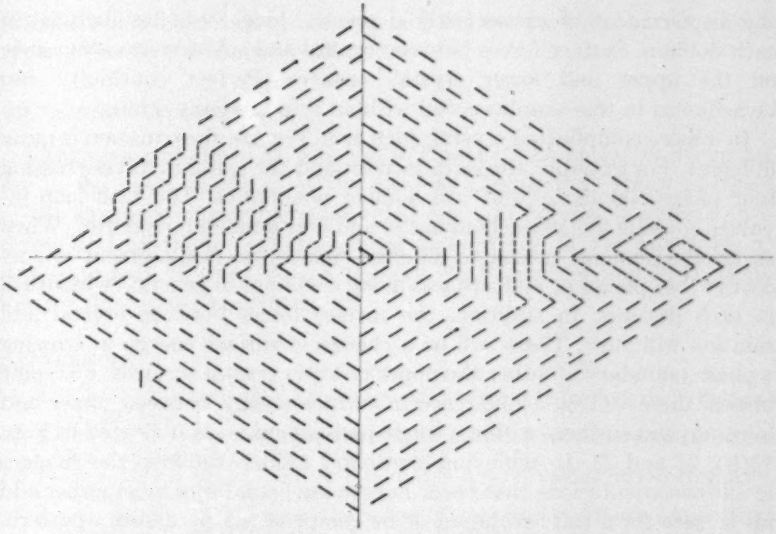


Fig. 29 — Hypothetical single-spiral crystal containing a variety of phase boundaries and domain boundaries, and showing ten molecules.

A pair of re-entrant faces are developed from the (200) fold domain. The inner corner formed by these faces would be expected to be a likely nucleation site. Suppose the experimental conditions are such that secondary nuclei of the monolayer type are favored. Then, to fill in the re-entrant area, a molecule would need to develop in a maze-like pattern such as is indicated in Fig. 29. However, this process should be very time-consuming because of the frequency with which corners must be rounded. Such a factor may offer a partial explanation for the widespread occurrence of re-entrant faces in many crystal aggregates.

VI. ELECTRON INTERFERENCE EFFECTS

The diversity and complexity of fold structures discussed in the preceding sections suggest an explanation for the types of electron interference effects which are illustrated in Figs. 5, 6 and 7. Whenever there is a change from one structure or domain to another, the associated crystal surface is likely to change its inclination to the (001) plane.

Certain restraints are placed on the morphology by the requirement of continuity across boundaries combined with a spiral growth process. The simplest possible loop around a spiral involves crossing four domain boundaries as in Fig. 24. Each domain is part of the same phase and

the displacement of successive fold planes along c can be identical in each domain. Surface forces between crystal and solution are symmetric on the upper and lower crystal surfaces. Perfect continuity can be achieved in this simple crystal without imposing any strain.

In a more complicated crystal, such as in Fig. 29, the situation is quite different. For example, the sixth loop around the spiral involves crossing four phase boundaries and two domain boundaries. The 12th loop involves crossing five phase boundaries and four domain boundaries. When successive domains belong to different phases the displacement of successive fold planes or fold surfaces along c will not in general be identical in each domain. In addition, the surface forces between crystal and solution will vary. There will be a change in surface energy in crossing a phase boundary at either the upper or lower crystal surfaces. For some phases there will be a difference in surface energy between upper and lower crystal surfaces within a single phase domain, as indicated in Figs. 13(b), 21 and 23. In achieving continuity around the loop the changes in surface orientations that occur in crossing boundaries must either add up to zero for a full revolution or be compensated by a strain pattern. As a complex structure develops, the crystal surfaces are likely to contain both discreet changes in orientation along well-defined crystallographic directions and continuous warpings or curvatures in both tangential and radial directions.

When a crystal is removed from solution and the solvent is allowed to evaporate still other changes may be expected. The distribution of surface tensions will be quite different for a crystal in air as opposed to those for a crystal in xylene. Adjustments in surface morphology are inevitable, and such effects as collapsed hollow pyramids and doubled-up lamellae may arise from such factors.

When electron interference effects give rise to contrast lines along well-defined directions, it is reasonable to suppose that in many instances they delineate phase boundaries or domain boundaries. When the contrast appears along curves or loops, continuous warpings or curvatures are being revealed. It is a moot point whether these nonlinear effects can be directly correlated with the position of boundaries. It is not unreasonable to think that they do indicate the occurrence, if not the actual position, of boundaries. The interpretation is bound to be complicated by extraneous deformations introduced in the process of drying and manipulating crystals for examination. With a limited number of experimental observations it is not reasonable to attempt more than a qualitative relation between interference effects and fold structure. With sufficient experimental data it becomes quite plausible that a definite cor-

relation could be obtained between a given contrast line and a specific type of phase or domain boundary.

VII. ACKNOWLEDGMENTS

I am very grateful to Miss S. Eloise Koonce, who patiently instructed me in the basic operation of the Siemens Elmiskop I. Many members of the chemistry department provided me with polymer materials, preprints and abstracts, encouragement and interest. I am especially indebted to J. B. Howard, J. L. Lundberg, W. P. Slichter and F. H. Winslow for their many courtesies.

REFERENCES

1. Till, P. H., The Growth of Single Crystals of Linear Polyethylene, J. Polymer Sci., **24,** 1957, p. 301.
2. Keller, A., A Note on Single Crystals in Polymers: Evidence for a Folded Chain Configuration, Phil. Mag., **2,** 1957, p. 1171.
3. Fischer, E. W., Stufen- und spiralförmiges Kristallwachstum bei Hochpolymeren, Z. Naturf., **12a,** 1957, p. 753.
4. Storks, K. H., An Electron Diffraction Examination of Some Linear High Polymers, J. Am. Chem. Soc., **60,** 1938, p. 1753.
5. Keller, A., and O'Connor, A., Study of Single Crystals and Their Associations in Polymers, Disc. Far. Soc., No. 25, 1958, p. 114.
6. Agar, A. W., Frank, F. C. and Keller, A., Crystallinity Effects in the Electron Microscopy of Polyethylene, Phil. Mag., **4,** 1959, p. 32.
7. Sella, C., and Trillat, J. J., Structures Périodiques dans les Polyéthylènes, C. R. Acad. Sci. Fr., **248,** 1959, p. 410.
8. Bassett, D. C., Frank, F. C., and Keller, A., Evidence for Distinct Sectors in Polymer Single Crystals, Nature, **184,** 1959, p. 810.
9. Niegisch, W. D., The Nucleation of Polyethylene Spherulites by Single Crystals, J. Polymer Sci., **40,** 1959, p. 263.
10. Niegisch, W. D. and Swan, P. R., Pyramidal Single Crystals of Polyethylene, Bull. Am. Phys. Soc., **5,** 1960, p. 192.
11. Reneker, D. H. and Geil, P. H., Morphology of Polymer Single Crystals, Bull. Am. Phys. Soc. **5,** 1960, p. 193.
12. Belavtseva, E. M., The Law of Mutual Substitution for the Process of Disordering of Some Crystal Compounds Under the Action of Moderate-Energy Electrons, Doklady Akad. Nauk SSSR, **125,** 1959, p. 1005.
13. Dowell, W. C. T., Farrant, J. L., and Rees, A. L. G., Electron Interference in Lamellar Crystals, Proc. Int. Conf. Electron Microscopy, London, 1954, p. 279.
14. Lauritzen, J. I., Jr. and Hoffman, J. D., Theory of Formation of Polymer Crystals with Folded Chains in Dilute Solution, J. Research, Natl. Bur. Standards, **64A,** 1960, p. 73.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

BAKANOWSKI, A. E., CRANNA, N. G. and UHLIR, A., JR.

Diffused Silicon Nonlinear Capacitors, Monograph 3588.

BARRY, P. H. and WHITMAN, A. L.

An Error-Detection System for 5-Unit-Code Teletypewriter Transmission, Monograph 3616.

BENEŠ, V. E.

Theory of Queues with One Server, Monograph 3612.

BÖMMEL, H. E. and DRANSFELD, K.

Excitation and Attenuation of Hypersonic Waves in Quartz, Monograph 3589.

BOWERS, K. D., see HEMPSTEAD, C. F.

CHYNOWETH, A. G.

Pyroelectricity, Internal Domains, and Interface Charges in Triglycine Sulfate, Monograph 3590.

CLOGSTON, A. M., GORDON, J. P., JACCARINO, V., PETER, M. and WALKER, L. R.

Hfs of F^{19} in the Electron Paramagnetic Resonance of Mn:ZnF_2, Monograph 3591.

CRANNA, N. G., see Bakanowski, A. E.

D'AMICO, C., see Hagstrum, H. D.

---

* Copies of these monographs may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. The numbers of the monographs should be given in all requests.

DE GRASSE, R. W., HOGG, D. C., OHM, E. A. and SCOVIL, H. E. D.

Ultra-Low Noise Receiving System for Satellite or Space Communication, Monograph 3624.

DE MONTE, R. W.

Synthesis of Cable Simulation Networks, Monograph 3487.

DEVERALL, G. V., see Purvis, M. B.

DODD, D. M.

Effects of X-Irradiation on Ferroelectric Triglycine Sulfate, Monograph 3614.

DODD, D. M.

Infrared Absorption Characteristics of Ferroelectric Triglycine Sulfate, Monograph 3617.

DRANSFELD, K., see Bömmel, H. E.

DRENICK, R. F.

Mathematical Aspects of the Reliability Problem, Monograph 3592.

DUNBAR, F. C.

A Delay Distortion Simulation Set, Monograph 3625.

DURAND, J. L., see Geller, S.

DURKEE, A. L., METCALFE, D. and TIDD, W. H.

White Alice System—Design and Performance, Monograph 3593.

FLETCHER, R. C., LE CRAW, R. C. and SPENCER, E. G.

Electron Spin Relaxation in Ferromagnetic Insulators, Monograph 3594.

FRISCH, H. L., see Rice, S. A.

GELLER, S.

Magnetic Interactions and Distribution of Ions in the Garnets, Monograph 3626.

GELLER, S. and DURAND, J. L.

Refinement of the Structure of $LiMnPO_4$, Monograph 3595.

GIGUERE, W. J., JAMISON, J. H. and NOLL, J. C.

Transistor Pulse Circuits for 160-MC Clock Rates, Monograph 3613.

GORDON, J. P., see Clogston, A. M.

GREEN, E. I.

Progress and Problems in World Communication, Monograph 3634.

GUPTA, S. S.

Order Statistics from the Gamma Distribution, Monograph 3597.

HAGSTRUM, H. D. and D'AMICO, C.

Production and Demonstration of Atomically Clean Metal Surfaces, Monograph 3598.

HEMPSTEAD, C. F. and BOWERS, K. D.

Parametric Resonance of Impurities in $CaWO_4$, Monograph 3599.

HOGG, D. C., see De Grasse, R. W.

JACCARINO, V., see Clogston, A. M.

JAMISON, J. H., see Giguere, W. J.

KLEINMAN, D. A. and SPITZER, W. G.

The GaP Crystal: Infrared Lattice Absorption; Anharmonic Forces, Monograph 3629.

LE CRAW, R. C., see Fletcher, R. C.

LUONGO, J. P.

Infrared Study of Oxygenated Groups Formed in Polyethylene during Oxidation, Monograph 3600.

MASON, W. P.

Phonon Viscosity and Its Effect on Acoustic Wave Attenuation and Dislocation, Monograph 3601.

MATTHIAS, B. T.

Relations Between Superconductors and Ferromagnets, Monograph 3630.

MAY, J. E., JR.

Thickness-Shear Mode $BaTiO_3$ Ceramic Transducers for Ultrasonic Delay Lines, Monograph 3602.

MCCLUSKEY, E. J., JR. and UNGER, S. H.

Number of Internal Variable Assignments for Sequential Switching Circuits, Monograph 3618.

METCALFE, D., see Durkee, A. L.

MILLER, R. C. and SAVAGE, A.

Domain-Walls in Metal Electroded Barium Titanate Crystals, Monograph 3603.

MILLER, R. C. and WEINREICH, G.

Mechanism for the Sidewise Motion of 180° Domain Walls in Barium Titanate, Monograph 3604.

MURPHY, R. B.

Stopping Rules with CSP-1 Sampling Inspection Plans, Monograph 2936.

NOLL, J. C., see Giguere, W. J.

OHM, E. A., see De Grasse, R. W.

OLSEN, K. M.

Effect of Trace Elements on High Purity Nickel, Monograph 3632.

PETER, M., see Clogston, A. M.

PURVIS, M. B. and DEVERALL, G. V.

Photographic Problems Associated with Construction of a Large Computer Memory, Monograph 3605.

RICE, S. A. and FRISCH, H. L.

Dynamical Theory of Diffusion in Crystals, Monograph 3606.

RUNGE, W. A., see Willson, F. E.

SAVAGE, A., see MILLER, R. C.

SCHEINMAN, A. H.

Numerical-Graphical Method in the Design of Multiterminal Switching Circuits, Monograph 3495.

SCHELKUNOFF, S. A.

Anatomy of "Surface Waves," Monograph 3615.

SCOVIL, H. E. D., see De Grasse, R. W.

SHAIR, R. C.

Lead-Acid Storage Batteries in Telephone Service, Monograph 3607.

SODEN, R. R., see Van Uitert, L. G.

SPENCER, E. G., see Fletcher, R. C.

SPITZER, W. G., see Kleinman, D. A.

THURSTON, R. N.

Electrical and Mechanical Terminating Resistances of a Piezoelectric Transducer, Monograph 3608.

TIDD, W. H., see Durkee, A. L.

UHLIR, A., JR., see BAKANOWSKI, A. E.

UNGER, S. H.

A Theorem on Self-Overlapping Sequences, Monograph 3611.

UNGER, S. H.

Pattern Detection and Recognition, Monograph 3609.

UNGER, S. H., see McCluskey, E. J., Jr.

VAN UITERT, L. G. and SODEN, R. R.

Emission Spectra of Trivalent Terbium, Monograph 3610.

WALKER, L. R., see Clogston, A. M.

WEINREICH, G., see Miller, R. C.

WHITMAN, A. L., see Barry, P. H.

WILLSON, F. E. and RUNGE, W. A.

Data Transmission Tests on Tropospheric Beyond-the-Horizon Radio System, Monograph 3619.

Contributors to This Issue

Arthur L. Bonner, B.S. in E.E., 1927, University of Minnesota; Columbia University, 1929–30; Northwestern Bell Telephone Co., 1919–23; Tri-State Telephone & Telegraph Co., 1924–27; Bell Telephone Laboratories, 1927—. As a member of the technical staff of Bell Laboratories, Mr. Bonner worked on step-by-step switching and has been primarily engaged in work on transmission development and voice-frequency, radio and carrier transmission systems.

C. Breen, E.E., 1930, Cornell University; American Telephone & Telegraph Co., 1930–38; Bell Telephone Laboratories, 1938—. After early assignments with the Long Lines Plant Department, Mr. Breen transferred to Bell Laboratories, where he was engaged in systems studies for No. 4 toll crossbar. During World War II he was concerned with military radar and microwave communication. After the war he was occupied with systems maintenance engineering on No. 4A and 5 crossbar, and later was in charge of work of this type on all local and toll switching, signaling and AMA systems. At present he is Special Systems Engineer in charge of work on PBXs and special service projects. Member Eta Kappa Nu.

Robinson D. Burbank, A.B., 1942, Colby College; Ph.D., 1950, Massachusetts Institute of Technology; Carbide & Carbon Chemicals Co., 1950–53; Olin-Mathieson Chemical Corp., 1953–55; Bell Telephone Laboratories, 1955—. He has been engaged in X-ray studies of soft magnetic alloys and in studies of methods of X-ray analysis. Member American Crystallographic Association, American Physical Society, Phi Beta Kappa, Sigma Xi.

T. R. D. Collins, A.B., 1925, John Fletcher College; M.S., 1929, State University of Iowa; Bell Telephone Laboratories, 1929—. His first work was on high-quality recording and reproducing systems and later development of coaxial cable. Following this he was concerned with television requirements and a television monitor system. During the war Mr. Collins worked on radar and a proximity fuse and since

the war he has been engaged in work on microwave radio relay system problems. Member American Physical Society.

HAROLD E. CURTIS, B.S. and M.S., 1929, Massachusetts Institute of Technology; American Telephone & Telegraph Co., 1929–34; Bell Telephone Laboratories, 1934—. He has specialized in work on transmission problems related to multi-channel carrier telephony, including microwave radio relay, coaxial cable and waveguide systems.

CARL A. DAHLBOM, B.E.E., 1941, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1930—. Mr. Dahlbom's first work was with the Telegraph Department in transmission and circuit development. During World War II he was concerned with development of military telegraph systems, transmission methods for speech-secrecy systems and tube modulators for radar systems. After the war he was engaged in design of single and multi-frequency signaling systems. He now heads a group responsible for formulation of signaling requirements for message service and transmission for telegraph systems. Senior member I.R.E.

MARK B. GARDNER, B.S., 1930, Brigham Young University; Bell Telephone Laboratories, 1930—. He has specialized in speech and hearing studies and problems in physiological acoustics, particularly as applied to individuals with impaired hearing. During World War II he worked on problems of underwater sound in connection with the development of the acoustic torpedo. Fellow Acoustical Society of America.

J. L. GARRISON, B.E.E., 1934, and M.E.E., 1936, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1936—. Mr. Garrison's first work was design of transmission transformers. He later was concerned with final development of transistors before turning to work on Bell Laboratories technical publications. In 1956 he transferred to Merrimack Valley, where he is now head of a group engaged in transmission networks development. Member Sigma Xi, Tau Beta Pi.

EUGENE I. GORDON, B.S., 1952, City College of New York; Ph.D., 1957, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1957—. A member of the Electron Device Department, he is engaged in research in beam-type parametric amplifiers. Member American Physical Society, Phi Beta Kappa, Sigma Xi.

PHILIP HUSTA, Western Electric Co., 1918–24; Bell Telephone Laboratories, 1924—. Mr. Husta's early work was in development of low-frequency signaling systems. He was later in charge of a group concerned with fundamental pulsing studies and development tests of switching circuits. During the war he worked on development of maintenance equipment for fire control equipment. Since the war he has been concerned with switching apparatus and he now heads a group concerned with capability and application of sealed-contact relays in switching systems. Member A.I.E.E.

B. C. JAMISON, B.S.E.E., 1930, Purdue University; American Telephone & Telegraph Co., Long Lines, 1927—. He was engaged in engineering carrier system layouts and in the problems resulting from the operation and maintenance of those systems. Since 1958 he has been doing similar work on microwave radio systems. He is in charge of a group in the General Methods Section of Long Lines Headquarters Engineering. Member A.I.E.E., Eta Kappa Nu, Tau Beta Pi.

WILLIAM J. KOPP, B.S. in E.E., 1937, New York University; Bell Telephone Laboratories, 1929—. Mr. Kopp first was concerned with transmission measuring apparatus related to studies on transmission standards and effective transmission rating systems. During World War II he worked on wire and radio networks and, after the war, on system aspects of early mobile radio telephone equipment. Since 1947 his work has been mainly concerned with negative impedance repeaters.

ANDREW D. LIEHR, B.S., 1952, University of Illinois; M.A., 1953 and Ph.D., 1955, Harvard University; National Science Foundation Post-Doctoral Fellow, Harvard University, 1955–56; Bell Telephone Laboratories, 1956—. He has been engaged in basic research in chemistry and physics, including studies of the electronic properties and structures of certain organic and inorganic materials with regard to their optical and magnetic properties. Member American Association for the Advancement of Science, American Association of Physics Teachers, American Chemical Society, American Institute of Physics, American Physical Society.

SAMUEL P. MORGAN, B.S., 1943, M.S., 1944, and Ph.D., 1947, California Institute of Technology; Bell Telephone Laboratories, 1947—. A research mathematician, Mr. Morgan has been particularly concerned with the applications of electromagnetic theory to microwave and other

problems. He now supervises a group doing research in various fields of mathematical physics. Senior Member I.R.E.; member American Physical Society, Sigma Xi, Tau Beta Pi.

GEORGE E. PERREAULT, B.S. in M.E., 1930, Worcester Polytechnic Institute; Bell Telephone Laboratories, 1930—. His early work was concerned with sound picture and picture transmission apparatus, and later magnetic tape recording and hearing-test machines for the Bell System exhibit in the New York World's Fair. During World War II he took part in work on electrical gun directors and other military projects. Since the war he has been engaged in work on selectors, switches and relays, and he now heads a group working on development of dry reed and mercury contact relays.

FRIEDOLF M. SMITS, Dipl. Phys., 1950, Dr. rer. nat., 1950, University of Freiburg, Germany; research associate, Physikalisches Institut, University of Freiburg, 1950–54; Bell Telephone Laboratories, 1954—. His past work included studies of solid-state diffusion in germanium and silicon, as well as device feasibility and process studies. At present he heads a group concerned with the development of UHF semiconductor devices. Member American Physical Society, German Physical Society.